DESCRIPTION: Diabetes is a serious disease affecting millions of people across the entire world. Thus, correct and timely prediction of this disease is very important due to the complications it can have in the case of other life-threatening diseases. High blood sugar level is the primary cause mostly seen in this disease. The objective of this project is to construct a prediction model for predicting diabetes using Pycaret. Pycaret, an open-source library consists of multiple classifiers and regressors for quickly selecting the best-performing algorithms. This allows you to prepare and deploy the model within minutes in your choice of notebook environment.

```
import matplotlib.pyplot as plt
%matplotlib inline
import pandas as pd
import seaborn as sns
import numpy as np
import plotly.offline as py
import plotly.express as px
import plotly.io as pio
import plotly.graph_objs as go
import math
from scipy.stats import norm, skew
import warnings
warnings.filterwarnings('ignore')
data = pd.read_csv("/content/Diabetes.csv")
data.head()
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

```
data.shape
```

```
(768, 9)
```

```
data = data.dropna()
data.isnull().sum()
```

```
Pregnancies                 0
Glucose                     0
BloodPressure               0
SkinThickness               0
Insulin                     0
BMI                         0
DiabetesPedigreeFunction    0
Age                         0
Outcome                     0
dtype: int64
```

```
data.info
```

```
<bound method DataFrame.info of      Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin   BMI  \
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
..           ...      ...            ...            ...      ...   ...
763           10      101             76             48      180  32.9
764            2      122             70             27        0  36.8
765            5      121             72             23      112  26.2
766            1      126             60              0        0  30.1
767            1       93             70             31        0  30.4

     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
..                        ...  ...      ...
763                     0.171   63        0
764                     0.340   27        0
765                     0.245   30        0
766                     0.349   47        1
767                     0.315   23        0

[768 rows x 9 columns]>
```

```
data.describe()
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	
count	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768
mean	3.845052	120.894531	69.105469	20.536458	79.799479	31.992578	0.471876	33
std	3.369578	31.972618	19.355807	15.952218	115.244002	7.884160	0.331329	11
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.078000	21
25%	1.000000	99.000000	62.000000	0.000000	0.000000	27.300000	0.243750	24
50%	3.000000	117.000000	72.000000	23.000000	30.500000	32.000000	0.372500	29
75%	6.000000	140.250000	80.000000	32.000000	127.250000	36.600000	0.626250	41
max	17.000000	199.000000	122.000000	99.000000	846.000000	67.100000	2.420000	81

```
data.fillna(data.mean(), inplace=True)
data.isnull().sum()
```

```
Pregnancies                 0
Glucose                     0
BloodPressure               0
SkinThickness               0
Insulin                     0
BMI                         0
DiabetesPedigreeFunction    0
Age                         0
Outcome                     0
dtype: int64
```

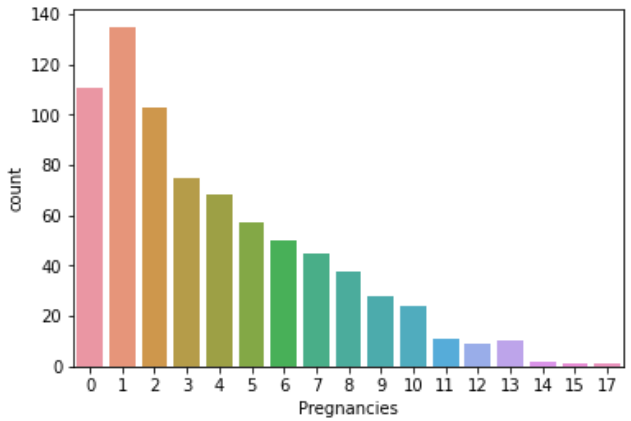
```
data.Pregnancies.value_counts()
```

```
1     135
0     111
2     103
3      75
4      68
5      57
6      50
7      45
8      38
9      28
10     24
11     11
13     10
12      9
```

```
14      2
15      1
17      1
Name: Pregnancies, dtype: int64
```

```
sns.countplot(data['Pregnancies'])
plt.show()
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variable as a
  FutureWarning
```

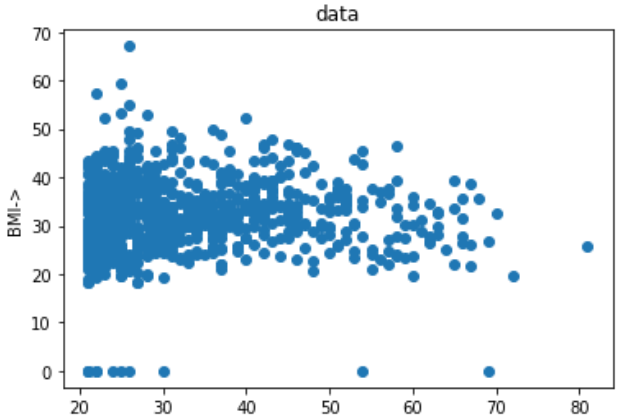



```
Age = data["Age"]
BMI = data["BMI"]
```

```
x=[]
y=[]
```

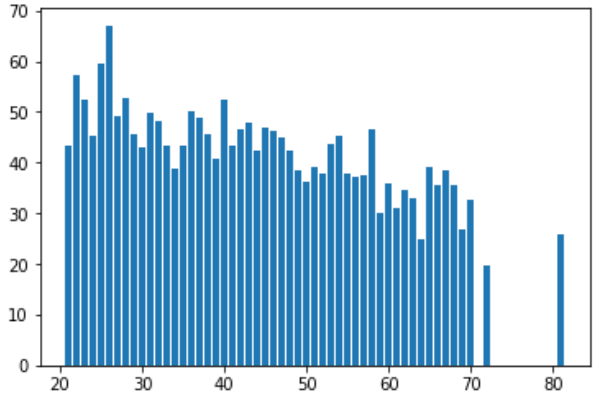
```
x=list(Age)
y=list(BMI)
```

```
plt.scatter(x,y)
plt.xlabel('Age->')
plt.ylabel('BMI->')
plt.title('data')
plt.show()
```




```
plt.bar(x,y)
```

```
<BarContainer object of 768 artists>
```

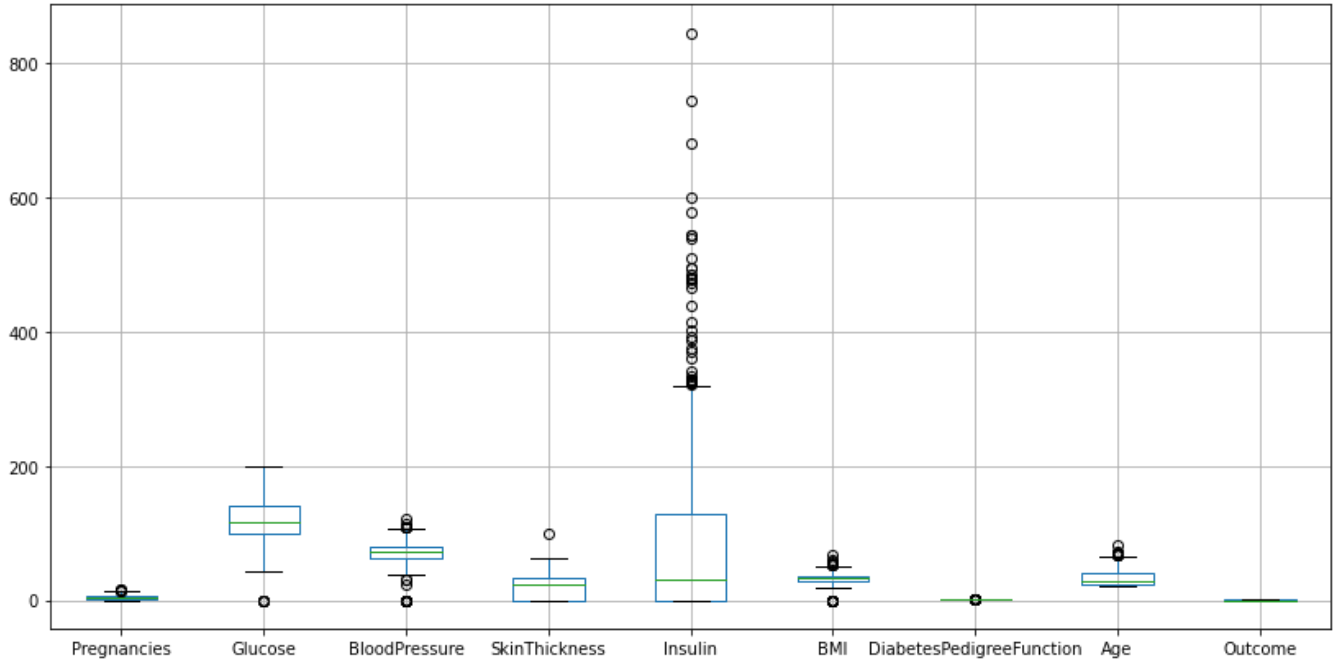



```
sns.countplot(data['Outcome'])
plt.show()
```

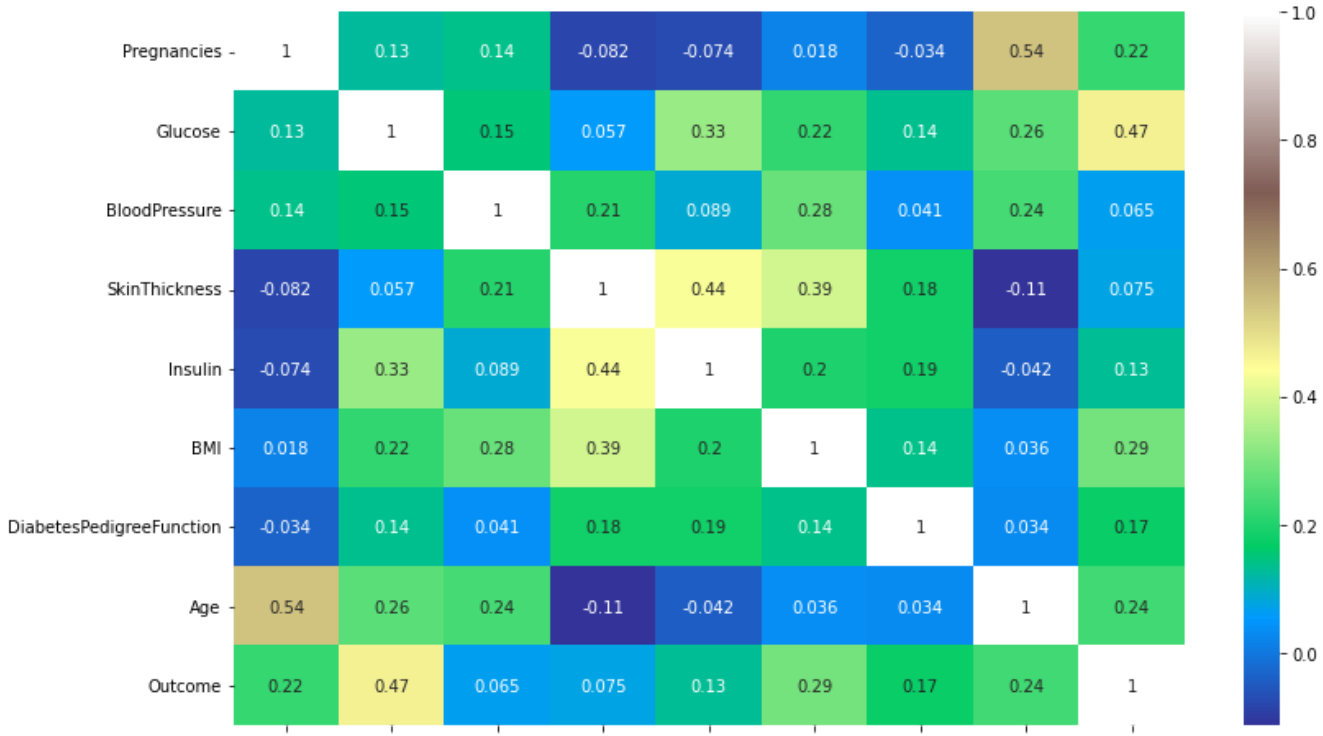
```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variable as a
  FutureWarning
```


```
data.boxplot(figsize=(14,7))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fd5f9a23bd0>
```




```
plt.figure(figsize=(13,8))
sns.heatmap(data.corr(),annot=True,cmap='terrain')
plt.show()
```



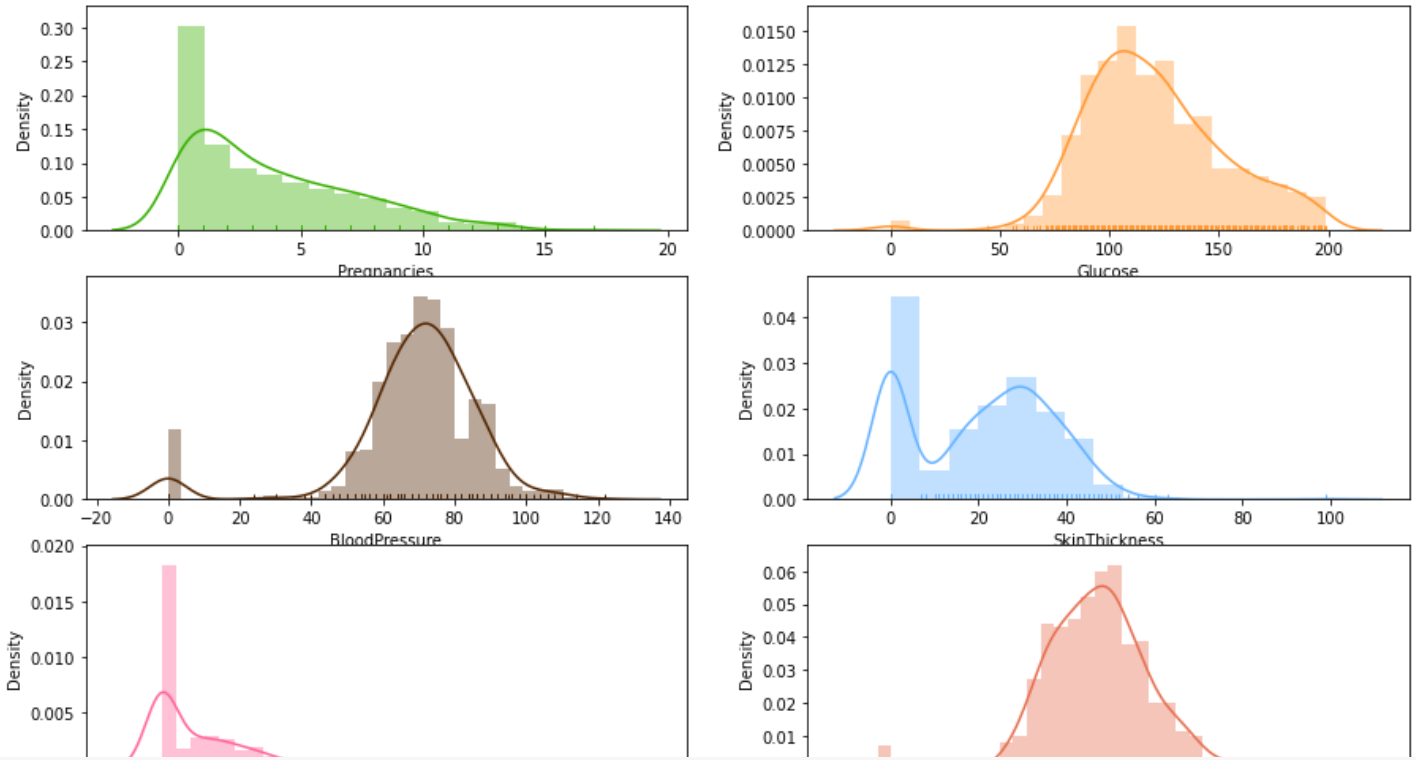

```
data.corr()
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigree
Pregnancies	1.000000	0.129459	0.141282	-0.081672	-0.073535	0.017683	
Glucose	0.129459	1.000000	0.152590	0.057328	0.331357	0.221071	
BloodPressure	0.141282	0.152590	1.000000	0.207371	0.088933	0.281805	
SkinThickness	-0.081672	0.057328	0.207371	1.000000	0.436783	0.392573	
Insulin	-0.073535	0.331357	0.088933	0.436783	1.000000	0.197859	
BMI	0.017683	0.221071	0.281805	0.392573	0.197859	1.000000	
DiabetesPedigreeFunction	-0.033523	0.137337	0.041265	0.183928	0.185071	0.140647	
Age	0.544341	0.263514	0.239528	-0.113970	-0.042163	0.036242	
Outcome	0.221898	0.466581	0.065068	0.074752	0.130548	0.292695	

```
data.info()
```

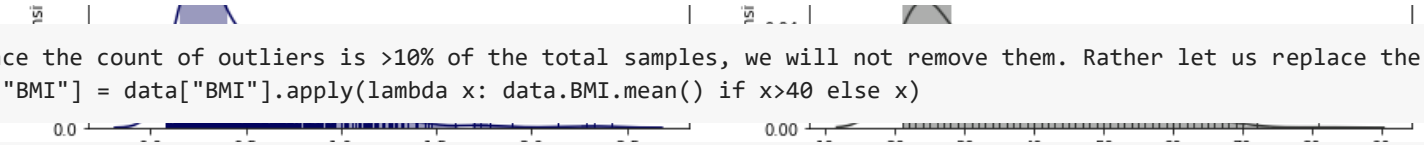
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   Pregnancies               768 non-null    int64  
 1   Glucose                   768 non-null    int64  
 2   BloodPressure             768 non-null    int64  
 3   SkinThickness             768 non-null    int64  
 4   Insulin                   768 non-null    int64  
 5   BMI                       768 non-null    float64
 6   DiabetesPedigreeFunction  768 non-null    float64
 7   Age                       768 non-null    int64  
 8   Outcome                   768 non-null    int64  
dtypes: float64(2), int64(7)
memory usage: 54.1 KB
```

```
# visulizing the 8 features
fig, axs = plt.subplots(4, 2, figsize=(15,12))
axs = axs.flatten()
sns.distplot(data['Pregnancies'],rug=True,color='#38b000',ax=axs[0])
sns.distplot(data['Glucose'],rug=True,color='#FF9933',ax=axs[1])
sns.distplot(data['BloodPressure'],rug=True,color='#522500',ax=axs[2])
sns.distplot(data['SkinThickness'],rug=True,color='#66b3ff',ax=axs[3])
sns.distplot(data['Insulin'],rug=True,color='#FF6699',ax=axs[4])
sns.distplot(data['BMI'],color='#e76f51',rug=True,ax=axs[5])
sns.distplot(data['DiabetesPedigreeFunction'],color='#03045e',rug=True,ax=axs[6])
sns.distplot(data['Age'],rug=True,color='#333533',ax=axs[7])
plt.show()
```




```
bmi_outliers=data[data['BMI']>40]
bmi_outliers['BMI'].shape
```

```
(96,)
```



```
# Since the count of outliers is >10% of the total samples, we will not remove them. Rather let us replace the BMI outliers (BMI>40) with the mean value
data["BMI"] = data["BMI"].apply(lambda x: data.BMI.mean() if x>40 else x)
```

```
# Setting up the model in PyCaret:
```

```
pip install pycaret[full]
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting pycaret[full]
  Downloading pycaret-2.3.10-py3-none-any.whl (320 kB)
     |████████████████████████████████| 320 kB 5.0 MB/s
Collecting pyyaml<6.0.0
  Downloading PyYAML-5.4.1-cp37-cp37m-manylinux1_x86_64.whl (636 kB)
     |████████████████████████████████| 636 kB 48.2 MB/s
Requirement already satisfied: yellowbrick>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from pycaret[full])
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (from pycaret[full]) (1.1.0)
Requirement already satisfied: ipywidgets in /usr/local/lib/python3.7/dist-packages (from pycaret[full]) (7.7.1)
Collecting scikit-learn==0.23.2
  Downloading scikit_learn-0.23.2-cp37-cp37m-manylinux1_x86_64.whl (6.8 MB)
     |████████████████████████████████| 6.8 MB 31.2 MB/s
Collecting mlflow
  Downloading mlflow-1.28.0-py3-none-any.whl (17.0 MB)
     |████████████████████████████████| 17.0 MB 36.4 MB/s
Collecting numba<0.55
  Downloading numba-0.54.1-cp37-cp37m-manylinux2014_x86_64.manylinux_2_17_x86_64.whl (3.3 MB)
     |████████████████████████████████| 3.3 MB 38.4 MB/s
Requirement already satisfied: plotly>=4.4.1 in /usr/local/lib/python3.7/dist-packages (from pycaret[full]) (5.5.
Collecting scipy<=1.5.4
  Downloading scipy-1.5.4-cp37-cp37m-manylinux1_x86_64.whl (25.9 MB)
     |████████████████████████████████| 25.9 MB 1.4 MB/s
Collecting lightgbm>=2.3.1
  Downloading lightgbm-3.3.2-py3-none-manylinux1_x86_64.whl (2.0 MB)
     |████████████████████████████████| 2.0 MB 41.5 MB/s
Requirement already satisfied: seaborn in /usr/local/lib/python3.7/dist-packages (from pycaret[full]) (0.11.2)
Requirement already satisfied: cufflinks>=0.17.0 in /usr/local/lib/python3.7/dist-packages (from pycaret[full]) (
Collecting kmodes>=0.10.1
  Downloading kmodes-0.12.2-py2.py3-none-any.whl (20 kB)
Collecting imbalanced-learn==0.7.0
  Downloading imbalanced_learn-0.7.0-py3-none-any.whl (167 kB)
     |████████████████████████████████| 167 kB 52.4 MB/s
Collecting pyLDAvis
  Downloading pyLDAvis-3.3.1.tar.gz (1.7 MB)
     |████████████████████████████████| 1.7 MB 59.3 MB/s
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
  Installing backend dependencies ... done
    Preparing wheel metadata ... done
Collecting pandas-profiling>=2.8.0
  Downloading pandas_profiling-3.3.0-py2.py3-none-any.whl (268 kB)
     |████████████████████████████████| 268 kB 55.8 MB/s
Collecting umap-learn
  Downloading umap-learn-0.5.3.tar.gz (88 kB)
     |████████████████████████████████| 88 kB 8.0 MB/s
Requirement already satisfied: IPython in /usr/local/lib/python3.7/dist-packages (from pycaret[full]) (7.9.0)
Requirement already satisfied: gensim<4.0.0 in /usr/local/lib/python3.7/dist-packages (from pycaret[full]) (3.6.0
Collecting spacy<2.4.0
  Downloading spacy-2.3.7-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (10.4 MB)
```

```
     |████████████████████████████████| 10.4 MB 35.2 MB/s
```

```
pip install pycaret
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: pycaret in /usr/local/lib/python3.7/dist-packages (2.3.10)
Requirement already satisfied: pyyaml<6.0.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (5.4.1)
Requirement already satisfied: nltk in /usr/local/lib/python3.7/dist-packages (from pycaret) (3.7)
Requirement already satisfied: plotly>=4.4.1 in /usr/local/lib/python3.7/dist-packages (from pycaret) (5.5.0)
Requirement already satisfied: pyod in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.0.4)
Requirement already satisfied: yellowbrick>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.3.post1)
Requirement already satisfied: kmodes>=0.10.1 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.12.2)
Requirement already satisfied: textblob in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.15.3)
Requirement already satisfied: pandas in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.3.5)
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.1.0)
Requirement already satisfied: scipy<=1.5.4 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.5.4)
Requirement already satisfied: pyLDAvis in /usr/local/lib/python3.7/dist-packages (from pycaret) (3.2.2)
Requirement already satisfied: wordcloud in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.8.2.2)
Requirement already satisfied: numba<0.55 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.54.1)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.7/dist-packages (from pycaret) (3.5.3)
Requirement already satisfied: scikit-learn==0.23.2 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.23.2)
Requirement already satisfied: IPython in /usr/local/lib/python3.7/dist-packages (from pycaret) (7.9.0)
Requirement already satisfied: seaborn in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.11.2)
Requirement already satisfied: imbalanced-learn==0.7.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.7.0)
Requirement already satisfied: pandas-profiling>=2.8.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (3.3.0)
Requirement already satisfied: ipywidgets in /usr/local/lib/python3.7/dist-packages (from pycaret) (7.7.1)
Requirement already satisfied: cufflinks>=0.17.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.17.3)
Requirement already satisfied: gensim<4.0.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (3.6.0)
Requirement already satisfied: scikit-plot in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.3.7)
Requirement already satisfied: mlxtend>=0.17.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.19.0)
Requirement already satisfied: mlflow in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.28.0)
Requirement already satisfied: umap-learn in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.5.3)
Requirement already satisfied: Boruta in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.3)
Requirement already satisfied: lightgbm>=2.3.1 in /usr/local/lib/python3.7/dist-packages (from pycaret) (3.3.2)
Requirement already satisfied: spacy<2.4.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (2.3.7)
Requirement already satisfied: numpy>=1.13.3 in /usr/local/lib/python3.7/dist-packages (from imbalanced-learn==0.7.0->pycaret) (1.19.5)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from scikit-learn==0.23.2->pycaret) (3.1.0)
Requirement already satisfied: colorlover>=0.2.1 in /usr/local/lib/python3.7/dist-packages (from cufflinks>=0.17.0->pycaret) (0.3.0)
Requirement already satisfied: six>=1.9.0 in /usr/local/lib/python3.7/dist-packages (from cufflinks>=0.17.0->pycaret) (1.15.0)
Requirement already satisfied: setuptools>=34.4.1 in /usr/local/lib/python3.7/dist-packages (from cufflinks>=0.17.0->pycaret) (57.4.0)
Requirement already satisfied: smart-open>=1.2.1 in /usr/local/lib/python3.7/dist-packages (from gensim<4.0.0->pycaret) (5.2.1)
Requirement already satisfied: pexpect in /usr/local/lib/python3.7/dist-packages (from IPython->pycaret) (4.8.0)
Requirement already satisfied: backcall in /usr/local/lib/python3.7/dist-packages (from IPython->pycaret) (0.2.0)
Requirement already satisfied: decorator in /usr/local/lib/python3.7/dist-packages (from IPython->pycaret) (4.4.2)
Requirement already satisfied: jedi>=0.10 in /usr/local/lib/python3.7/dist-packages (from IPython->pycaret) (0.18.1)
Requirement already satisfied: pickleshare in /usr/local/lib/python3.7/dist-packages (from IPython->pycaret) (0.7.5)
Requirement already satisfied: traitlets>=4.2 in /usr/local/lib/python3.7/dist-packages (from IPython->pycaret) (5.1.1)
Requirement already satisfied: prompt-toolkit<2.1.0,>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from IPython->pycaret) (2.0.10)
Requirement already satisfied: pygments in /usr/local/lib/python3.7/dist-packages (from IPython->pycaret) (2.6.1)
Requirement already satisfied: ipykernel>=4.5.1 in /usr/local/lib/python3.7/dist-packages (from ipywidgets->pycaret) (5.3.4)
```

```
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets->pycaret) (3.0.2)
Requirement already satisfied: widgetsnbextension~=3.6.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets->pycaret) (3.6.1)
Requirement already satisfied: ipython-genutils~=0.2.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets->pycaret) (0.2.0)
Requirement already satisfied: tornado>=4.2 in /usr/local/lib/python3.7/dist-packages (from ipykernel>=4.5.1->ipywidgets->pycaret) (5.1.1)
Requirement already satisfied: jupyter-client in /usr/local/lib/python3.7/dist-packages (from ipykernel>=4.5.1->ipywidgets->pycaret) (6.1.12)
Requirement already satisfied: parso<0.9.0,>=0.8.0 in /usr/local/lib/python3.7/dist-packages (from jedi>=0.10->IPython->pycaret) (0.8.3)
Requirement already satisfied: wheel in /usr/local/lib/python3.7/dist-packages (from lightgbm>=2.3.1->pycaret) (0.37.1)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages (from matplotlib->pycaret) (0.11.0)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.7/dist-packages (from matplotlib->pycaret) (2.8.2)
Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.7/dist-packages (from matplotlib->pycaret) (7.1.2)
```

```
pip install --pre pycaret
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting pycaret
  Downloading pycaret-3.0.0rc3-py3-none-any.whl (544 kB)
     |████████████████████████████████| 544 kB 8.2 MB/s
Collecting numba~=0.55.0
  Downloading numba-0.55.2-cp37-cp37m-manylinux2014_x86_64.manylinux_2_17_x86_64.whl (3.3 MB)
     |████████████████████████████████| 3.3 MB 50.1 MB/s
Collecting pmdarima>=1.8.0
  Downloading pmdarima-2.0.1-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.manylinux_2_28_x86_64.whl (1.8 MB)
     |████████████████████████████████| 1.8 MB 51.7 MB/s
Collecting schemdraw>=0.14
  Downloading schemdraw-0.15-py3-none-any.whl (106 kB)
     |████████████████████████████████| 106 kB 64.2 MB/s
Requirement already satisfied: imbalanced-learn>=0.8.1 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.8.1)
Requirement already satisfied: numpy~=1.21 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.21.6)
Requirement already satisfied: jinja2>=1.2 in /usr/local/lib/python3.7/dist-packages (from pycaret) (2.11.3)
Requirement already satisfied: markupsafe>=2.0.1 in /usr/local/lib/python3.7/dist-packages (from pycaret) (2.0.1)
Collecting matplotlib>=3.3.0
  Downloading matplotlib-3.5.3-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86_64.whl (11.2 MB)
     |████████████████████████████████| 11.2 MB 45.1 MB/s
Requirement already satisfied: yellowbrick>=1.4 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.5)
Collecting sktime~=0.11.4
  Downloading sktime-0.11.4-py3-none-any.whl (6.7 MB)
     |████████████████████████████████| 6.7 MB 62.5 MB/s
Requirement already satisfied: scikit-learn>=1.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.0.2)
Requirement already satisfied: scipy<1.9.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.7.3)
Collecting kaleido>=0.2.1
  Downloading kaleido-0.2.1-py2.py3-none-manylinux1_x86_64.whl (79.9 MB)
     |████████████████████████████████| 79.9 MB 113 kB/s
Requirement already satisfied: pandas<1.5.0,>=1.3.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.3.5)
Requirement already satisfied: joblib>=1.1.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.1.0)
Requirement already satisfied: statsmodels>=0.12.1 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.12.2)
Collecting category-encoders>=2.4.0
  Downloading category_encoders-2.5.0-py2.py3-none-any.whl (69 kB)
     |████████████████████████████████| 69 kB 7.9 MB/s
Requirement already satisfied: ipywidgets>=7.6.5 in /usr/local/lib/python3.7/dist-packages (from pycaret) (7.7.1)
Requirement already satisfied: plotly>=5.0.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (5.5.0)
Requirement already satisfied: ipython>=5.5.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (7.9.0)
Collecting requests>=2.27.1
  Downloading requests-2.28.1-py3-none-any.whl (62 kB)
     |████████████████████████████████| 62 kB 1.4 MB/s
Collecting tbats>=1.1.0
  Downloading tbats-1.1.0-py3-none-any.whl (43 kB)
     |████████████████████████████████| 43 kB 2.5 MB/s
Collecting lightgbm>=3.0.0
  Downloading lightgbm-3.3.2-py3-none-manylinux1_x86_64.whl (2.0 MB)
     |████████████████████████████████| 2.0 MB 10.7 MB/s
Collecting plotly-resampler>=0.7.2.2
  Downloading plotly-resampler-0.8.2rc1.tar.gz (94 kB)
     |████████████████████████████████| 94 kB 2.2 MB/s
```

```
  Installing build dependencies ... done
```

```
from pycaret.datasets import get_data
diabetes = get_data('diabetes')
```

	Number of times pregnant	Plasma glucose concentration a 2 hours in an oral glucose tolerance test	Diastolic blood pressure (mm Hg)	Triceps skin fold thickness (mm)	2-Hour serum insulin (mu U/ml)	Body mass index (weight in kg/(height in m)^2)	Diabetes pedigree function	Age (years)	Class variable
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0

```
  Collecting jedi>=0.10
```

```
type(diabetes)
```

```
pandas.core.frame.DataFrame
```

```
Requirement already satisfied: setuptools>=18.5 in /usr/local/lib/python3.7/dist-packages (from ipython>=5.5.0->pycaret) (57
```

```
diabetes.head()
```

	Number of times pregnant	Plasma glucose concentration a 2 hours in an oral glucose tolerance test	Diastolic blood pressure (mm Hg)	Triceps skin fold thickness (mm)	2-Hour serum insulin (mu U/ml)	Body mass index (weight in kg/(height in m)^2)	Diabetes pedigree function	Age (years)	Class variable
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0

```
Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.7/dist-packages (from matplotlib>=3.3.0->pycaret) (7.1
```

```
diabetes.dtypes
```

```
Number of times pregnant                                                  int64
Plasma glucose concentration a 2 hours in an oral glucose tolerance test  int64
Diastolic blood pressure (mm Hg)                                          int64
Triceps skin fold thickness (mm)                                          int64
2-Hour serum insulin (mu U/ml)                                            int64
Body mass index (weight in kg/(height in m)^2)                          float64
```

```
Diabetes pedigree function                                              float64
Age (years)                                                               int64
Class variable                                                            int64
dtype: object
```

```
  Downloading dash-2.6.1-py3-none-any.whl (9.9 MB)
```

```
# setting up Environment
from pycaret.classification import *
```

```
     |████████████████████████████████| 270 kB 55.6 MB/s
```

```
# Intializing the Setup
exp_clf = setup(diabetes, target = 'Class variable')
#exp_clf = setup(diabetes, target = 'Class variable', silent=True)
```

```
INFO:logs:PyCaret ClassificationExperiment
INFO:logs:Logging name: clf-default-name
INFO:logs:ML Usecase: MLUsecase.CLASSIFICATION
INFO:logs:version 3.0.0.rc3
INFO:logs:Initializing setup()
INFO:logs:self.USI: 1452
INFO:logs:self.variable_keys: {'exp_name_log', 'USI', 'y_train', '_ml_usecase', '_is_multiclass', '_gpu_n_jobs_pa
INFO:logs:Checking environment
INFO:logs:python_version: 3.7.13
INFO:logs:python_build: ('default', 'Apr 24 2022 01:04:09')
INFO:logs:machine: x86_64
INFO:logs:platform: Linux-5.10.133+-x86_64-with-Ubuntu-18.04-bionic
INFO:logs:Memory: svmem(total=13616361472, available=11638042624, percent=14.5, used=1863041024, free=8877580288,
INFO:logs:Physical Core: 1
INFO:logs:Logical Core: 2
INFO:logs:Checking libraries
INFO:logs:System:
INFO:logs:    python: 3.7.13 (default, Apr 24 2022, 01:04:09)  [GCC 7.5.0]
INFO:logs:executable: /usr/bin/python3
INFO:logs:   machine: Linux-5.10.133+-x86_64-with-Ubuntu-18.04-bionic
INFO:logs:PyCaret required dependencies:
INFO:logs:                 pip: 21.1.3
INFO:logs:          setuptools: 57.4.0
INFO:logs:             pycaret: 3.0.0.rc3
INFO:logs:             IPython: 7.9.0
INFO:logs:          ipywidgets: 7.7.1
INFO:logs:                tqdm: 4.64.0
INFO:logs:               numpy: 1.21.6
INFO:logs:              pandas: 1.3.5
INFO:logs:              jinja2: 2.11.3
INFO:logs:               scipy: 1.7.3
INFO:logs:              joblib: 1.1.0
INFO:logs:             sklearn: 1.0.2
INFO:logs:                pyod: Installed but version unavailable
INFO:logs:            imblearn: 0.8.1
INFO:logs:   category_encoders: 2.5.0
INFO:logs:            lightgbm: 3.3.2
INFO:logs:               numba: 0.55.2
INFO:logs:            requests: 2.28.1
INFO:logs:          matplotlib: 3.2.2
INFO:logs:          scikitplot: 0.3.7
INFO:logs:         yellowbrick: 1.5
INFO:logs:              plotly: 5.5.0
INFO:logs:             kaleido: 0.2.1
INFO:logs:         statsmodels: 0.13.2
INFO:logs:              sktime: 0.11.4
INFO:logs:               tbats: Installed but version unavailable
INFO:logs:            pmdarima: 2.0.1
INFO:logs:              psutil: 5.4.8
INFO:logs:None
```

```
INFO:logs:Set up data.
```

```
# Data Type interference
#exp_clf = setup(diabetes, target = 'Class variable' ,categorical_features=['Age (years)'])
exp_clf = setup(diabetes, target = 'Class variable' ,numeric_features=['Age (years)'])
#exp_clf = setup(diabetes, target = 'Class variable' ,ignore_features=['Age (years)'])
#exp_clf = setup(diabetes, target = 'Class variable' ,date_features=['Age (years)'])
```

```
INFO:logs:PyCaret ClassificationExperiment
INFO:logs:Logging name: clf-default-name
INFO:logs:ML Usecase: MLUsecase.CLASSIFICATION
INFO:logs:version 3.0.0.rc3
INFO:logs:Initializing setup()
INFO:logs:self.USI: f546
INFO:logs:self.variable_keys: {'exp_name_log', 'USI', 'y_train', '_ml_usecase', '_is_multiclass', '_gpu_n_jobs_pa
INFO:logs:Checking environment
INFO:logs:python_version: 3.7.13
INFO:logs:python_build: ('default', 'Apr 24 2022 01:04:09')
INFO:logs:machine: x86_64
INFO:logs:platform: Linux-5.10.133+-x86_64-with-Ubuntu-18.04-bionic
INFO:logs:Memory: svmem(total=13616361472, available=11529691136, percent=15.3, used=1974820864, free=8750653440,
INFO:logs:Physical Core: 1
INFO:logs:Logical Core: 2
INFO:logs:Checking libraries
INFO:logs:System:
INFO:logs:    python: 3.7.13 (default, Apr 24 2022, 01:04:09)  [GCC 7.5.0]
INFO:logs:executable: /usr/bin/python3
INFO:logs:   machine: Linux-5.10.133+-x86_64-with-Ubuntu-18.04-bionic
INFO:logs:PyCaret required dependencies:
INFO:logs:                 pip: 21.1.3
INFO:logs:          setuptools: 57.4.0
INFO:logs:             pycaret: 3.0.0.rc3
INFO:logs:             IPython: 7.9.0
INFO:logs:          ipywidgets: 7.7.1
INFO:logs:                tqdm: 4.64.0
INFO:logs:               numpy: 1.21.6
INFO:logs:              pandas: 1.3.5
INFO:logs:              jinja2: 2.11.3
INFO:logs:               scipy: 1.7.3
INFO:logs:              joblib: 1.1.0
INFO:logs:             sklearn: 1.0.2
INFO:logs:                pyod: Installed but version unavailable
INFO:logs:            imblearn: 0.8.1
INFO:logs:   category_encoders: 2.5.0
INFO:logs:            lightgbm: 3.3.2
INFO:logs:               numba: 0.55.2
INFO:logs:            requests: 2.28.1
INFO:logs:          matplotlib: 3.2.2
INFO:logs:          scikitplot: 0.3.7
INFO:logs:         yellowbrick: 1.5
INFO:logs:              plotly: 5.5.0
INFO:logs:             kaleido: 0.2.1
INFO:logs:         statsmodels: 0.13.2
INFO:logs:              sktime: 0.11.4
INFO:logs:               tbats: Installed but version unavailable
INFO:logs:            pmdarima: 2.0.1
INFO:logs:              psutil: 5.4.8
INFO:logs:None
```

```
INFO:logs:Set up data.
```

```
# Data Cleaning And Preparation
exp_clf = setup(diabetes, target = 'Class variable' ,numeric_imputation= 'median')
#exp_clf = setup(diabetes, target = 'Class variable' ,categorical_imputation== 'mode')
```

```
INFO:logs:PyCaret ClassificationExperiment
INFO:logs:Logging name: clf-default-name
INFO:logs:ML Usecase: MLUsecase.CLASSIFICATION
INFO:logs:version 3.0.0.rc3
INFO:logs:Initializing setup()
INFO:logs:self.USI: f035
INFO:logs:self.variable_keys: {'exp_name_log', 'USI', 'y_train', '_ml_usecase', '_is_multiclass', '_gpu_n_jobs_pa
INFO:logs:Checking environment
INFO:logs:python_version: 3.7.13
INFO:logs:python_build: ('default', 'Apr 24 2022 01:04:09')
INFO:logs:machine: x86_64
INFO:logs:platform: Linux-5.10.133+-x86_64-with-Ubuntu-18.04-bionic
INFO:logs:Memory: svmem(total=13616361472, available=11514744832, percent=15.4, used=1989869568, free=8728428544,
INFO:logs:Physical Core: 1
INFO:logs:Logical Core: 2
INFO:logs:Checking libraries
INFO:logs:System:
INFO:logs:    python: 3.7.13 (default, Apr 24 2022, 01:04:09)  [GCC 7.5.0]
INFO:logs:executable: /usr/bin/python3
INFO:logs:   machine: Linux-5.10.133+-x86_64-with-Ubuntu-18.04-bionic
INFO:logs:PyCaret required dependencies:
INFO:logs:                 pip: 21.1.3
INFO:logs:          setuptools: 57.4.0
INFO:logs:             pycaret: 3.0.0.rc3
INFO:logs:             IPython: 7.9.0
INFO:logs:          ipywidgets: 7.7.1
INFO:logs:                tqdm: 4.64.0
INFO:logs:               numpy: 1.21.6
INFO:logs:              pandas: 1.3.5
INFO:logs:              jinja2: 2.11.3
INFO:logs:               scipy: 1.7.3
INFO:logs:              joblib: 1.1.0
INFO:logs:             sklearn: 1.0.2
INFO:logs:                pyod: Installed but version unavailable
INFO:logs:            imblearn: 0.8.1
INFO:logs:   category_encoders: 2.5.0
INFO:logs:            lightgbm: 3.3.2
INFO:logs:               numba: 0.55.2
INFO:logs:            requests: 2.28.1
INFO:logs:          matplotlib: 3.2.2
INFO:logs:          scikitplot: 0.3.7
INFO:logs:         yellowbrick: 1.5
INFO:logs:              plotly: 5.5.0
INFO:logs:             kaleido: 0.2.1
INFO:logs:         statsmodels: 0.13.2
INFO:logs:              sktime: 0.11.4
INFO:logs:               tbats: Installed but version unavailable
INFO:logs:            pmdarima: 2.0.1
INFO:logs:              psutil: 5.4.8
INFO:logs:None
```

```
INFO:logs:Set up data.
```

```
compare_models()
```

```
INFO:logs:Initializing compare_models()
INFO:logs:compare_models(self=<pycaret.classification.oop.ClassificationExperiment object at 0x7efcf0066590>, inc
INFO:logs:Checking exceptions
INFO:logs:Preparing display monitor
```

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
lda	Linear Discriminant Analysis	0.7823	0.8380	0.5939	0.7308	0.6515	0.4967	0.5049	0.0390
ridge	Ridge Classifier	0.7785	0.0000	0.5775	0.7310	0.6402	0.4848	0.4950	0.0320
lr	Logistic Regression	0.7784	0.8375	0.5936	0.7245	0.6472	0.4892	0.4982	0.0680
et	Extra Trees Classifier	0.7656	0.8433	0.5520	0.7090	0.6175	0.4537	0.4628	0.2670
qda	Quadratic Discriminant Analysis	0.7581	0.8295	0.5991	0.6771	0.6285	0.4515	0.4588	0.0390
rf	Random Forest Classifier	0.7579	0.8368	0.5567	0.6925	0.6115	0.4403	0.4495	0.2990
nb	Naive Bayes	0.7543	0.8303	0.6105	0.6606	0.6285	0.4469	0.4527	0.0390
gbc	Gradient Boosting Classifier	0.7487	0.8358	0.5836	0.6551	0.6132	0.4294	0.4336	0.1310
lightgbm	Light Gradient Boosting Machine	0.7263	0.8153	0.5626	0.6198	0.5867	0.3839	0.3871	0.0700
ada	Ada Boost Classifier	0.7245	0.7945	0.5515	0.6137	0.5783	0.3757	0.3783	0.1170
dt	Decision Tree Classifier	0.7095	0.6778	0.5728	0.5843	0.5769	0.3565	0.3574	0.0390
knn	K Neighbors Classifier	0.7077	0.7270	0.4705	0.6117	0.5248	0.3206	0.3304	0.1380
dummy	Dummy Classifier	0.6518	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0350
svm	SVM - Linear Kernel	0.6148	0.0000	0.4462	0.5216	0.3958	0.1542	0.1865	0.0310

Processing: 100% 61/61 [00:18<00:00, 4.76it/s]

```
INFO:logs:Initializing Logistic Regression
INFO:logs:Total runtime is 4.976987838745117e-05 minutes
INFO:logs:SubProcess create_model() called ==================================
INFO:logs:Initializing create_model()
INFO:logs:create_model(self=<pycaret.classification.oop.ClassificationExperiment object at 0x7efcf0066590>, estim
INFO:logs:Checking exceptions
INFO:logs:Importing libraries
INFO:logs:Copying training dataset
INFO:logs:Defining folds
INFO:logs:Declaring metric variables
INFO:logs:Importing untrained model
INFO:logs:Logistic Regression Imported successfully
INFO:logs:Starting cross validation
INFO:logs:Cross validating with StratifiedKFold(n_splits=10, random_state=None, shuffle=False), n_jobs=-1
INFO:logs:Calculating mean and std
INFO:logs:Creating metrics dataframe
INFO:logs:Uploading results into container
```

```
INFO:logs:Uploading model into container now
```

```
best_model = compare_models()
```

```
INFO:logs:Initializing compare_models()
INFO:logs:compare_models(self=<pycaret.classification.oop.ClassificationExperiment object at 0x7efcf0066590>, inc
INFO:logs:Checking exceptions
INFO:logs:Preparing display monitor
```

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
lda	Linear Discriminant Analysis	0.7823	0.8380	0.5939	0.7308	0.6515	0.4967	0.5049	0.0400
ridge	Ridge Classifier	0.7785	0.0000	0.5775	0.7310	0.6402	0.4848	0.4950	0.0310
lr	Logistic Regression	0.7784	0.8375	0.5936	0.7245	0.6472	0.4892	0.4982	0.5200
et	Extra Trees Classifier	0.7656	0.8433	0.5520	0.7090	0.6175	0.4537	0.4628	0.2680
qda	Quadratic Discriminant Analysis	0.7581	0.8295	0.5991	0.6771	0.6285	0.4515	0.4588	0.0380
rf	Random Forest Classifier	0.7579	0.8368	0.5567	0.6925	0.6115	0.4403	0.4495	0.3090
nb	Naive Bayes	0.7543	0.8303	0.6105	0.6606	0.6285	0.4469	0.4527	0.0380
gbc	Gradient Boosting Classifier	0.7487	0.8358	0.5836	0.6551	0.6132	0.4294	0.4336	0.1320
lightgbm	Light Gradient Boosting Machine	0.7263	0.8153	0.5626	0.6198	0.5867	0.3839	0.3871	0.1400
ada	Ada Boost Classifier	0.7245	0.7945	0.5515	0.6137	0.5783	0.3757	0.3783	0.1150
dt	Decision Tree Classifier	0.7095	0.6778	0.5728	0.5843	0.5769	0.3565	0.3574	0.0390
knn	K Neighbors Classifier	0.7077	0.7270	0.4705	0.6117	0.5248	0.3206	0.3304	0.1380
dummy	Dummy Classifier	0.6518	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0350
svm	SVM - Linear Kernel	0.6148	0.0000	0.4462	0.5216	0.3958	0.1542	0.1865	0.0330

Processing: 100% 61/61 [00:23<00:00, 4.11it/s]

```
INFO:logs:Initializing Logistic Regression
INFO:logs:Total runtime is 5.100568135579427e-05 minutes
INFO:logs:SubProcess create_model() called ==================================
INFO:logs:Initializing create_model()
INFO:logs:create_model(self=<pycaret.classification.oop.ClassificationExperiment object at 0x7efcf0066590>, estim
INFO:logs:Checking exceptions
INFO:logs:Importing libraries
INFO:logs:Copying training dataset
INFO:logs:Defining folds
INFO:logs:Declaring metric variables
INFO:logs:Importing untrained model
INFO:logs:Logistic Regression Imported successfully
INFO:logs:Starting cross validation
INFO:logs:Cross validating with StratifiedKFold(n_splits=10, random_state=None, shuffle=False), n_jobs=-1
INFO:logs:Calculating mean and std
INFO:logs:Creating metrics dataframe
INFO:logs:Uploading results into container
```

```
INFO:logs:Uploading model into container now
```

```
rf = create_model('rf', fold = 10)
```

```
INFO:logs:Initializing create_model()
INFO:logs:create_model(self=<pycaret.classification.oop.ClassificationExperiment object at 0x7efcf0066590>, estim
INFO:logs:Checking exceptions
```

Fold	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	
0	0.7222	0.8301	0.4211	0.6667	0.5161	0.3350	0.3524	.....

```
tuned_rf = tune_model(rf)
```

```
INFO:logs:Initializing tune_model()
INFO:logs:tune_model(estimator=RandomForestClassifier(bootstrap=True, ccp_alpha=0.0, class_weight=None,
                       criterion='gini', max_depth=None, max_features='auto',
                       max_leaf_nodes=None, max_samples=None,
                       min_impurity_decrease=0.0, min_samples_leaf=1,
                       min_samples_split=2, min_weight_fraction_leaf=0.0,
                       n_estimators=100, n_jobs=-1, oob_score=False,
                       random_state=1936, verbose=0, warm_start=False), fold=None, round=4, n_iter=10, custom_gri
INFO:logs:Checking exceptions
```

Fold	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.7778	0.8361	0.6316	0.7059	0.6667	0.5008	0.5025
1	0.8333	0.8917	0.7895	0.7500	0.7692	0.6389	0.6394
2	0.7963	0.8812	0.7368	0.7000	0.7179	0.5587	0.5591
3	0.7037	0.7955	0.6842	0.5652	0.6190	0.3802	0.3848
4	0.6296	0.7248	0.5263	0.4762	0.5000	0.2070	0.2077
5	0.7407	0.8150	0.6842	0.6190	0.6500	0.4449	0.4463
6	0.8519	0.8902	0.8947	0.7391	0.8095	0.6901	0.6985
7	0.7925	0.8778	0.8333	0.6522	0.7317	0.5665	0.5779
8	0.6981	0.7905	0.5556	0.5556	0.5556	0.3270	0.3270
9	0.8302	0.9048	0.7778	0.7368	0.7568	0.6265	0.6270
Mean	0.7654	0.8408	0.7114	0.6500	0.6776	0.4941	0.4970
Std	0.0675	0.0555	0.1123	0.0885	0.0935	0.1456	0.1470

```
Processing: 100%                                                    7/7 [00:33<00:00, 3.85s/it]
INFO:logs:Copying training dataset
INFO:logs:Checking base model
INFO:logs:Base model : Random Forest Classifier
INFO:logs:Declaring metric variables
INFO:logs:Defining Hyperparameters
INFO:logs:Tuning with n_jobs=-1
INFO:logs:Initializing RandomizedSearchCV
Fitting 10 folds for each of 10 candidates, totalling 100 fits
INFO:logs:best_params: {'actual_estimator__n_estimators': 220, 'actual_estimator__min_samples_split': 2, 'actual_
INFO:logs:Hyperparameter search completed
INFO:logs:SubProcess create_model() called ==================================
INFO:logs:Initializing create_model()
INFO:logs:create_model(self=<pycaret.classification.oop.ClassificationExperiment object at 0x7efcf0066590>, estim
                       criterion='gini', max_depth=None, max_features='auto',
```

```
                          max_leaf_nodes=None, max_samples=None
```

```
tuned_rf
```

```
RandomForestClassifier(bootstrap=True, ccp_alpha=0.0, class_weight='balanced',
                       criterion='entropy', max_depth=6, max_features='sqrt',
                       max_leaf_nodes=None, max_samples=None,
                       min_impurity_decrease=0.002, min_samples_leaf=2,
                       min_samples_split=2, min_weight_fraction_leaf=0.0,
                       n_estimators=220, n_jobs=-1, oob_score=False,
                       random_state=1936, verbose=0, warm_start=False)
```

```
INFO:logs:Declaring custom model
```

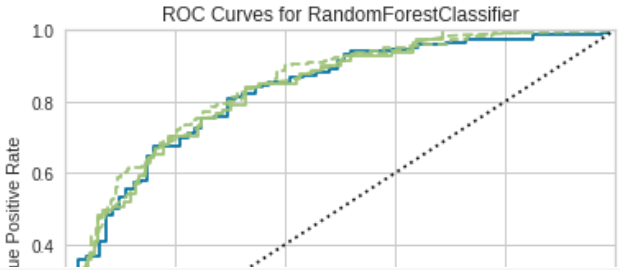
```
pip install matplotlib==3.1.3
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting matplotlib==3.1.3
  Downloading matplotlib-3.1.3-cp37-cp37m-manylinux1_x86_64.whl (13.1 MB)
     |████████████████████████████████| 13.1 MB 6.1 MB/s
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages (from matplotlib==3.1.3) (0
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib==3.1.
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib==3
Requirement already satisfied: numpy>=1.11 in /usr/local/lib/python3.7/dist-packages (from matplotlib==3.1.3) (1.
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.7/dist-packages (from kiwisolver>=1.0.
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (from python-dateutil>=2.1->mat
Installing collected packages: matplotlib
  Attempting uninstall: matplotlib
    Found existing installation: matplotlib 3.5.3
    Uninstalling matplotlib-3.5.3:
      Successfully uninstalled matplotlib-3.5.3
ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This b
pycaret 3.0.0rc3 requires matplotlib>=3.3.0, but you have matplotlib 3.1.3 which is incompatible.
Successfully installed matplotlib-3.1.3
```

```
                          max_leaf_nodes=None, max_samples=None,
```

```
plot_model(tuned_rf, plot = 'auc')
```

```
INFO:logs:Initializing plot_model()
INFO:logs:plot_model(plot=auc, fold=None, use_train_data=False, verbose=True, display=None, display_format=None,
                     criterion='entropy', max_depth=6, max_features='sqrt',
                     max_leaf_nodes=None, max_samples=None,
                     min_impurity_decrease=0.002, min_samples_leaf=2,
                     min_samples_split=2, min_weight_fraction_leaf=0.0,
                     n_estimators=220, n_jobs=-1, oob_score=False,
                     random_state=1936, verbose=0, warm_start=False), feature_name=None, fit_kwargs=None, group
INFO:logs:Checking exceptions
INFO:logs:Preloading libraries
INFO:logs:Copying training dataset
INFO:logs:Plot type: auc
INFO:logs:Fitting Model
INFO:logs:Scoring test/hold-out set
```



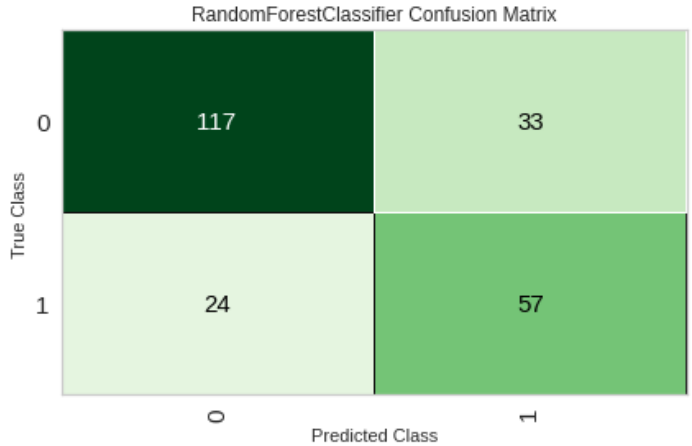

```
, warm_start=False) result for Accuracy is 0.7579
```

```
plot_model(tuned_rf, plot = 'pr')
```

```
INFO:logs:Initializing plot_model()
INFO:logs:plot_model(plot=pr, fold=None, use_train_data=False, verbose=True, display=None, display_format=None, e
                        criterion='entropy', max_depth=6, max_features='sqrt',
                        max_leaf_nodes=None, max_samples=None,
                        min_impurity_decrease=0.002, min_samples_leaf=2,
                        min_samples_split=2, min_weight_fraction_leaf=0.0,
                        n_estimators=220, n_jobs=-1, oob_score=False,
                        random_state=1936, verbose=0, warm_start=False), feature_name=None, fit_kwargs=None, group
INFO:logs:Checking exceptions
INFO:logs:Preloading libraries
INFO:logs:Importing libraries
```

```
plot_model(tuned_rf, plot = 'confusion_matrix')
```

```
INFO:logs:Initializing plot_model()
INFO:logs:plot_model(plot=confusion_matrix, fold=None, use_train_data=False, verbose=True, display=None, display_
                        criterion='entropy', max_depth=6, max_features='sqrt',
                        max_leaf_nodes=None, max_samples=None,
                        min_impurity_decrease=0.002, min_samples_leaf=2,
                        min_samples_split=2, min_weight_fraction_leaf=0.0,
                        n_estimators=220, n_jobs=-1, oob_score=False,
                        random_state=1936, verbose=0, warm_start=False), feature_name=None, fit_kwargs=None, group
INFO:logs:Checking exceptions
INFO:logs:Preloading libraries
INFO:logs:Copying training dataset
INFO:logs:Plot type: confusion_matrix
INFO:logs:Fitting Model
INFO:logs:Scoring test/hold-out set
```




```
INFO:logs:Visual Rendered Successfully
INFO:logs:plot_model() successfully completed.....................................

INFO:logs:Uploading model into container now
```

```
!pip install matplotlib-venn
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: matplotlib-venn in /usr/local/lib/python3.7/dist-packages (0.11.7)
Requirement already satisfied: numpy in /usr/local/lib/python3.7/dist-packages (from matplotlib-venn) (1.21.6)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.7/dist-packages (from matplotlib-venn) (3.1.3)
Requirement already satisfied: scipy in /usr/local/lib/python3.7/dist-packages (from matplotlib-venn) (1.7.3)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib->matplotlib-venn) (1.4.4)
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib->matplotlib-venn) (2.8.2)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages (from matplotlib->matplotlib-venn) (0.11.0)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib->matplotlib-venn) (3.0.9)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.7/dist-packages (from kiwisolver>=1.0.1->matplotlib->matplotlib-venn) (4.1.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (from python-dateutil>=2.1->matplotlib->matplotlib-venn) (1.15.0)
INFO:logs:SubProcess create_model() called ==================================
```

```
!apt-get -qq install -y libfluidsynth1
```

```
Selecting previously unselected package libfluidsynth1:amd64.
(Reading database ... 155685 files and directories currently installed.)
Preparing to unpack .../libfluidsynth1_1.1.9-1_amd64.deb ...
Unpacking libfluidsynth1:amd64 (1.1.9-1) ...
Setting up libfluidsynth1:amd64 (1.1.9-1) ...
Processing triggers for libc-bin (2.27-3ubuntu1.5) ...
INFO:logs:Starting cross validation
```

```
# https://pypi.python.org/pypi/libarchive
!apt-get -qq install -y libarchive-dev && pip install -U libarchive
import libarchive
```

```
Selecting previously unselected package libarchive-dev:amd64.
(Reading database ... 155690 files and directories currently installed.)
Preparing to unpack .../libarchive-dev_3.2.2-3.1ubuntu0.7_amd64.deb ...
Unpacking libarchive-dev:amd64 (3.2.2-3.1ubuntu0.7) ...
Setting up libarchive-dev:amd64 (3.2.2-3.1ubuntu0.7) ...
Processing triggers for man-db (2.8.3-2ubuntu0.1) ...
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting libarchive
  Downloading libarchive-0.4.7.tar.gz (23 kB)
Collecting nose
  Downloading nose-1.3.7-py3-none-any.whl (154 kB)
     |████████████████████████████████| 154 kB 8.3 MB/s
Building wheels for collected packages: libarchive
  Building wheel for libarchive (setup.py) ... done
  Created wheel for libarchive: filename=libarchive-0.4.7-py3-none-any.whl size=31646 sha256=742a11b31715a421c32cbf4fdd31373d71cccc462244e3372ca148c2e6c9ae92
  Stored in directory: /root/.cache/pip/wheels/63/b1/c6/b3da79bec2012175bd43603eed98ef8548ac1733b77c1d4330
Successfully built libarchive
Installing collected packages: nose, libarchive
Successfully installed libarchive-0.4.7 nose-1.3.7
INFO:logs:Declaring metric variables
```

```
!apt-get -qq install -y graphviz && pip install pydot
import pydot
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: pydot in /usr/local/lib/python3.7/dist-packages (1.3.0)
Requirement already satisfied: pyparsing>=2.1.4 in /usr/local/lib/python3.7/dist-packages (from pydot) (3.0.9)
```

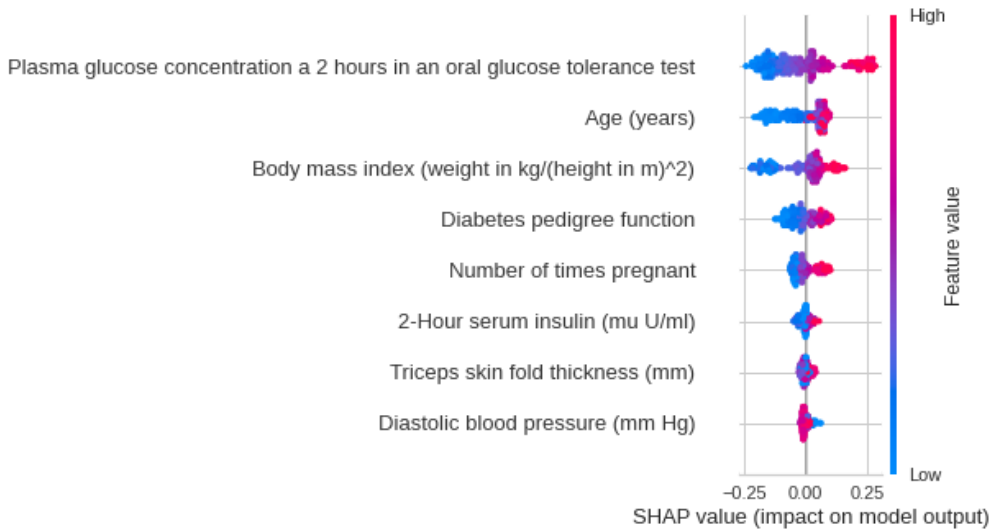
```
!pip install cartopy
import cartopy
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting cartopy
  Downloading Cartopy-0.20.3.tar.gz (10.8 MB)
     |████████████████████████████████| 10.8 MB 4.9 MB/s
  Installing build dependencies ... done
  Getting requirements to build wheel ... error
WARNING: Discarding https://files.pythonhosted.org/packages/98/a9/0e4000eabadfcff6373c0fec790863b543b919cbfec18aed60d71ba67d5d/Cartopy-0.20.3.tar.gz#sha256=0
  Downloading Cartopy-0.20.2.tar.gz (10.8 MB)
     |████████████████████████████████| 10.8 MB 18.3 MB/s
  Installing build dependencies ... done
  Getting requirements to build wheel ... error
WARNING: Discarding https://files.pythonhosted.org/packages/f6/55/1e1c737dc9436b320deead73d1c455ddbb74b8b6992081863492f6f6378a/Cartopy-0.20.2.tar.gz#sha256=4
  Downloading Cartopy-0.20.1.tar.gz (10.8 MB)
     |████████████████████████████████| 10.8 MB 23.6 MB/s
  Installing build dependencies ... done
  Getting requirements to build wheel ... error
WARNING: Discarding https://files.pythonhosted.org/packages/fc/59/aa52698e3838f4cd0e7eaa75bd86837e9e0b05041dbdaee3cda2fffced06/Cartopy-0.20.1.tar.gz#sha256=91
  Downloading Cartopy-0.20.0.tar.gz (10.8 MB)
     |████████████████████████████████| 10.8 MB 22.2 MB/s
  Installing build dependencies ... done
  Getting requirements to build wheel ... error
WARNING: Discarding https://files.pythonhosted.org/packages/0f/c0/58453b036e79046d211f083880d58dcce787e7e07647ac25dc46c6555099/Cartopy-0.20.0.tar.gz#sha256=ea
  Downloading Cartopy-0.19.0.post1.tar.gz (12.1 MB)
     |████████████████████████████████| 12.1 MB 21.1 MB/s
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
    Preparing wheel metadata ... done
Collecting pyshp>=2
  Downloading pyshp-2.3.1-py2.py3-none-any.whl (46 kB)
     |████████████████████████████████| 46 kB 4.6 MB/s
Requirement already satisfied: numpy>=1.13.3 in /usr/local/lib/python3.7/dist-packages (from cartopy) (1.21.6)
Requirement already satisfied: shapely>=1.5.6 in /usr/local/lib/python3.7/dist-packages (from cartopy) (1.8.4)
Building wheels for collected packages: cartopy
  Building wheel for cartopy (PEP 517) ... done
  Created wheel for cartopy: filename=Cartopy-0.19.0.post1-cp37-cp37m-linux_x86_64.whl size=12516294 sha256=af358aa23962764aa9538c1532e4d31ee2f0520009ab7a2413
  Stored in directory: /root/.cache/pip/wheels/98/01/f7/bd10aeb96fe4b518cde5f7c4f5e12c7202f85b7353a5017847
Successfully built cartopy
```

```
Installing collected packages: pyshp, cartopy
Successfully installed cartopy-0.19.0.post1 pyshp-2.3.1
```

```
interpret_model(tuned_rf)
```

```
INFO:logs:Initializing interpret_model()
INFO:logs:interpret_model(estimator=RandomForestClassifier(bootstrap=True, ccp_alpha=0.0, class_weight='balanced'
                        criterion='entropy', max_depth=6, max_features='sqrt',
                        max_leaf_nodes=None, max_samples=None,
                        min_impurity_decrease=0.002, min_samples_leaf=2,
                        min_samples_split=2, min_weight_fraction_leaf=0.0,
                        n_estimators=220, n_jobs=-1, oob_score=False,
                        random_state=1936, verbose=0, warm_start=False), use_train_data=False, X_new_sample=None,
INFO:logs:Checking exceptions
INFO:logs:Soft dependency imported: shap: 0.41.0
INFO:logs:plot type: summary
INFO:logs:Creating TreeExplainer
INFO:logs:Compiling shap values
```




```
INFO:logs:Visual Rendered Successfully
INFO:logs:interpret_model() succesfully completed.....................................
```

```
predict_model(tuned_rf);
```

```
INFO:logs:Initializing predict_model()
INFO:logs:predict_model(self=<pycaret.classification.oop.ClassificationExperiment object at 0x7efcf0066590>, esti
                        criterion='entropy', max_depth=6, max_features='sqrt',
                        max_leaf_nodes=None, max_samples=None,
                        min_impurity_decrease=0.002, min_samples_leaf=2,
                        min_samples_split=2, min_weight_fraction_leaf=0.0,
                        n_estimators=220, n_jobs=-1, oob_score=False,
                        random_state=1936, verbose=0, warm_start=False), probability_threshold=None, encoded_label
INFO:logs:Checking exceptions
INFO:logs:Preloading libraries
```

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC

```
print(best_model)
```

```
GradientBoostingClassifier(ccp_alpha=0.0, criterion='friedman_mse', init=None,
                           learning_rate=0.1, loss='deviance', max_depth=3,
                           max_features=None, max_leaf_nodes=None,
                           min_impurity_decrease=0.0, min_samples_leaf=1,
                           min_samples_split=2, min_weight_fraction_leaf=0.0,
                           n_estimators=100, n_iter_no_change=None,
                           random_state=123, subsample=1.0, tol=0.0001,
                           validation_fraction=0.1, verbose=0,
                           warm_start=False)
```

```
# Shap Library
!pip install shap
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting shap
  Downloading shap-0.41.0-cp37-cp37m-manylinux_2_12_x86_64.manylinux2010_x86_64.whl (569 kB)
     |████████████████████████████████| 569 kB 6.7 MB/s
Collecting slicer==0.0.7
  Downloading slicer-0.0.7-py3-none-any.whl (14 kB)
Requirement already satisfied: cloudpickle in /usr/local/lib/python3.7/dist-packages (from shap) (1.5.0)
Requirement already satisfied: numba in /usr/local/lib/python3.7/dist-packages (from shap) (0.55.2)
Requirement already satisfied: packaging>20.9 in /usr/local/lib/python3.7/dist-packages (from shap) (21.3)
Requirement already satisfied: tqdm>4.25.0 in /usr/local/lib/python3.7/dist-packages (from shap) (4.64.0)
Requirement already satisfied: numpy in /usr/local/lib/python3.7/dist-packages (from shap) (1.21.6)
Requirement already satisfied: scipy in /usr/local/lib/python3.7/dist-packages (from shap) (1.7.3)
Requirement already satisfied: pandas in /usr/local/lib/python3.7/dist-packages (from shap) (1.3.5)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.7/dist-packages (from shap) (1.0.2)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /usr/local/lib/python3.7/dist-packages (from packaging>20.9->shap) (3.0.9)
Requirement already satisfied: llvmlite<0.39,>=0.38.0rc1 in /usr/local/lib/python3.7/dist-packages (from numba->shap) (0.38.1)
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages (from numba->shap) (57.4.0)
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/dist-packages (from pandas->shap) (2.8.2)
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-packages (from pandas->shap) (2022.2.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (from python-dateutil>=2.7.3->pandas->shap) (1.15.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from scikit-learn->shap) (3.1.0)
```

```
Requirement already satisfied: joblib>=0.11 in /usr/local/lib/python3.7/dist-packages (from scikit-learn->shap) (1.1.0)
Installing collected packages: slicer, shap
Successfully installed shap-0.41.0 slicer-0.0.7
```

```
import sklearn
from sklearn.model_selection import train_test_split
import numpy as np
import shap
import time

X,y = shap.datasets.diabetes()
X_train,X_test,y_train,y_test = train_test_split(X, y, test_size=0.2, random_state=0)
# rather than use the whole training set to estimate expected values, we summarzire
# aset of weighted k means, each weighted by the number of points they represent.
# X_train_summary = shap.kmeans(X_train, 10)
```

```
X
```

	age	sex	bmi	bp	s1	s2	s3	s4	s5	s6
0	0.038076	0.050680	0.061696	0.021872	-0.044223	-0.034821	-0.043401	-0.002592	0.019908	-0.017646
1	-0.001882	-0.044642	-0.051474	-0.026328	-0.008449	-0.019163	0.074412	-0.039493	-0.068330	-0.092204
2	0.085299	0.050680	0.044451	-0.005671	-0.045599	-0.034194	-0.032356	-0.002592	0.002864	-0.025930
3	-0.089063	-0.044642	-0.011595	-0.036656	0.012191	0.024991	-0.036038	0.034309	0.022692	-0.009362
4	0.005383	-0.044642	-0.036385	0.021872	0.003935	0.015596	0.008142	-0.002592	-0.031991	-0.046641
...	...	...	...	...	...	...	...	...	...	...
437	0.041708	0.050680	0.019662	0.059744	-0.005697	-0.002566	-0.028674	-0.002592	0.031193	0.007207
438	-0.005515	0.050680	-0.015906	-0.067642	0.049341	0.079165	-0.028674	0.034309	-0.018118	0.044485
439	0.041708	0.050680	-0.015906	0.017282	-0.037344	-0.013840	-0.024993	-0.011080	-0.046879	0.015491
440	-0.045472	-0.044642	0.039062	0.001215	0.016318	0.015283	-0.028674	0.026560	0.044528	-0.025930
441	-0.045472	-0.044642	-0.073030	-0.081414	0.083740	0.027809	0.173816	-0.039493	-0.004220	0.003064

442 rows × 10 columns

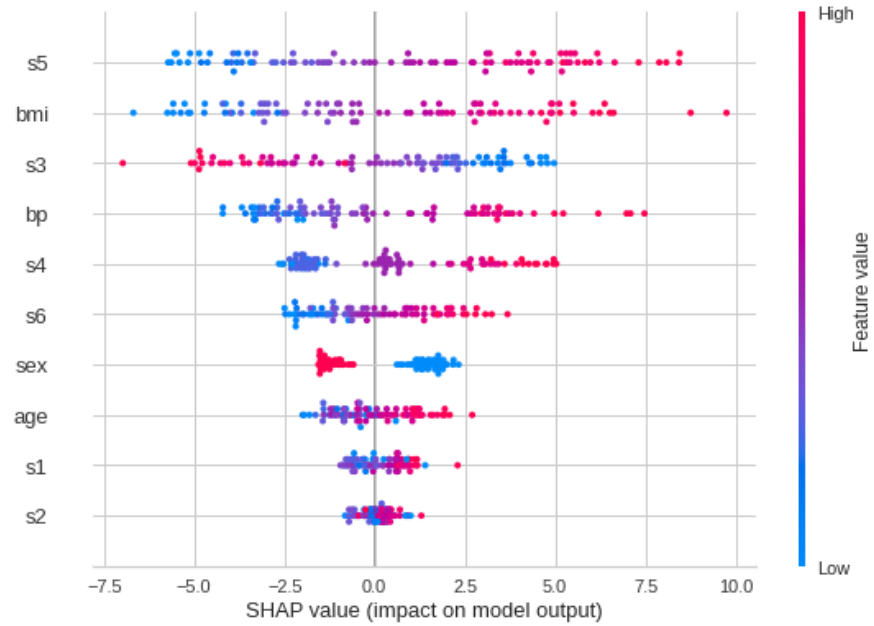
```
import sklearn
model = sklearn.svm.SVR()
model.fit(X_train, y_train)
```

```
SVR(C=1.0, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
    kernel='rbf', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
```

```
ex = shap.KernelExplainer(model.predict, shap.sample(X_train,100))
```

```
shap_values = ex.shap_values(X_test)
shap.summary_plot(shap_values, X_test)
```

```
100%                                                  89/89 [04:11<00:00, 2.78s/it]
```




```
shap_values.shape
```

```
(89, 10)
```

```
X_test.shape
```

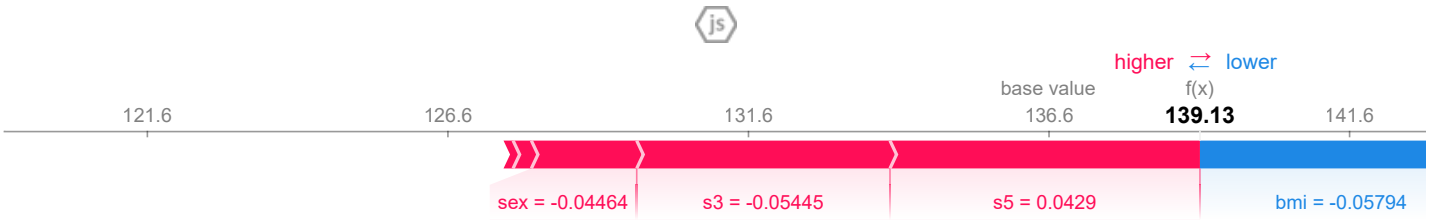
```
(89, 10)
```

```
print(model.predict(X_test.iloc[:11]))
```

```
[157.0600018  162.28980284 142.4484795  126.54597614 138.7299094
 159.14963825 117.93423588 154.71585068 133.25456956 152.3716958
 139.13151104]
```

```
from numpy.lib import shape_base
```

```
# particular value say for 10Th value i.e 139.13151104
shap.initjs()
shap_values = ex.shap_values(X_test.iloc[10,:])
shap.force_plot(ex.expected_value, shap_values, X_test.iloc[10,:])
```



Colab paid products - Cancel contracts here