MAJOR PROJECT – 1

INTRODUCTION

NAME – DRISHTI SINGH

BRANCH – COMPUTER SCIENCE AND ENGINEERING

COLLEGE – CHANDIGARH UNIVERSITY

YEAR – 2ND

AIM - Choose any dataset of your choice and apply a suitable CLASSIFIER/REGRESSOR

SCREENSHOTS-

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

```
Requirement already satisfied: tbats>=1.1.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.1.0)
Requirement already satisfied: scikit-learn>=1.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.0.2)
Requirement already satisfied: imbalanced-learn>=0.8.1 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.8.1)
Requirement already satisfied: importlib-metadata>=4.12.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (4.12.0)
Requirement already satisfied: pyod>=0.9.8 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.0.5)
Requirement already satisfied: category-encoders>=2.4.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (2.5.0)
Requirement already satisfied: numba~=0.55.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.55.2)
Requirement already satisfied: joblib>=1.1.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.1.0)
Requirement already satisfied: requests>=2.27.1 in /usr/local/lib/python3.7/dist-packages (from pycaret) (2.28.1)
Requirement already satisfied: numpy<1.23,>=1.21 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.21.6)
Requirement already satisfied: scikit-plot>=0.3.7 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.3.7)
Requirement already satisfied: statsmodels>=0.12.1 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.12.2)
Requirement already satisfied: pmdarima!=1.8.1,<2.0.0,>=1.8.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.8.5)
Requirement already satisfied: scipy<1.9.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.7.3)
Requirement already satisfied: tqdm>=4.62.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (4.64.1)
Requirement already satisfied: sktime~=0.13.2 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.13.4)
Requirement already satisfied: jinja2>=1.2 in /usr/local/lib/python3.7/dist-packages (from pycaret) (2.11.3)
Requirement already satisfied: matplotlib>=3.3.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (3.5.3)
Requirement already satisfied: lightgbm>=3.0.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (3.3.2)
Requirement already satisfied: plotly>=5.0.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (5.5.0)
Requirement already satisfied: plotly-resampler>=0.7.2.2 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.8.2rc3)
Requirement already satisfied: yellowbrick>=1.4 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.5)
```

```
print ("Installed sucessfully!!")
```

```
Installed sucessfully!!
```

```
[ ] from pycaret.utils import version
    version()
```

```
'3.0.0.rc4'
```

```
[ ] from pycaret.datasets import get_data
```

```
[ ] dataSets = get_data('index')
```

26	asia_gdp	Multivariate	Clustering	None	None	40	11	N
27	elections	Multivariate	Clustering	None	None	3195	54	Y
28	facebook	Multivariate	Clustering	None	None	7050	12	N
29	ipl	Multivariate	Clustering	None	None	153	25	N
30	jewellery	Multivariate	Clustering	None	None	505	4	N
31	mice	Multivariate	Clustering	None	None	1080	82	Y
32	migration	Multivariate	Clustering	None	None	233	12	N
33	perfume	Multivariate	Clustering	None	None	20	29	N
34	pokemon	Multivariate	Clustering	None	None	800	13	Y
35	population	Multivariate	Clustering	None	None	255	56	Y
36	public_health	Multivariate	Clustering	None	None	224	21	N

```
[ ] DataSet = get_data("blood")
```

	Recency	Frequency	Monetary	Time	Class
0	2	50	12500	98	1
1	0	13	3250	28	1
2	1	16	4000	35	1
3	2	20	5000	45	1
4	1	24	6000	77	0

```
[ ] from pycaret.datasets import get_data
    from pycaret.classification import *

    DataSet = get_data("blood")
    s = setup(data=DataSet, target='Class')

    sModel = create_model('svm')
```

2	Target type	Binary
3	Original data shape	(748, 5)
4	Transformed data shape	(748, 5)
5	Transformed train set shape	(523, 5)
6	Transformed test set shape	(225, 5)
7	Numeric features	4
8	Preprocess	True
9	Imputation type	simple
10	Numeric imputation	mean

[ ]

Fold	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.7547	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.7547	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.7547	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.7692	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.5577	0.0000	0.9167	0.3333	0.4889	0.2274	0.3208
5	0.7692	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.7692	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.7692	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.3462	0.0000	1.0000	0.2609	0.4138	0.0753	0.1978
9	0.4615	0.0000	1.0000	0.3171	0.4815	0.1642	0.2990
Mean	0.6706	0.0000	0.2917	0.0911	0.1384	0.0467	0.0818
Std	0.1489	0.0000	0.4460	0.1402	0.2122	0.0791	0.1283

[ ]
```
sm = save_model(sModel, 'sModelFile')
```

```
Transformation Pipeline and Model Successfully Saved
```

[ ]
```
sModel = load_model('sModelFile')
```

```
Transformation Pipeline and Model Successfully Loaded
```

[ ]
```
# Select top 10 rows from diabetes dataset
newDataSet = get_data("blood").iloc[:10]
```

	Recency	Frequency	Monetary	Time	Class
0	2	50	12500	98	1
1	0	13	3250	28	1
2	1	16	4000	35	1
3	2	20	5000	45	1

```
[ ]  newPredictions = predict_model(sModel, data = newDataSet)
     newPredictions
```

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	SVM - Linear Kernel	0.7000	0.5000	1.0000	0.7000	0.8235	0.0000	0.0000

	Recency	Frequency	Monetary	Time	Class	prediction_label
0	2.0	50.0	12500.0	98.0	1	1
1	0.0	13.0	3250.0	28.0	1	1
2	1.0	16.0	4000.0	35.0	1	1
3	2.0	20.0	5000.0	45.0	1	1
4	1.0	24.0	6000.0	77.0	0	1
5	4.0	4.0	1000.0	4.0	0	1
6	2.0	7.0	1750.0	14.0	1	1
7	1.0	12.0	3000.0	35.0	0	1
8	2.0	9.0	2250.0	22.0	1	1

```
[ ]  newPredictions.to_csv("NewPredictions.csv")
     # No output
```

```
[ ]  rfModel = create_model('svm')
```

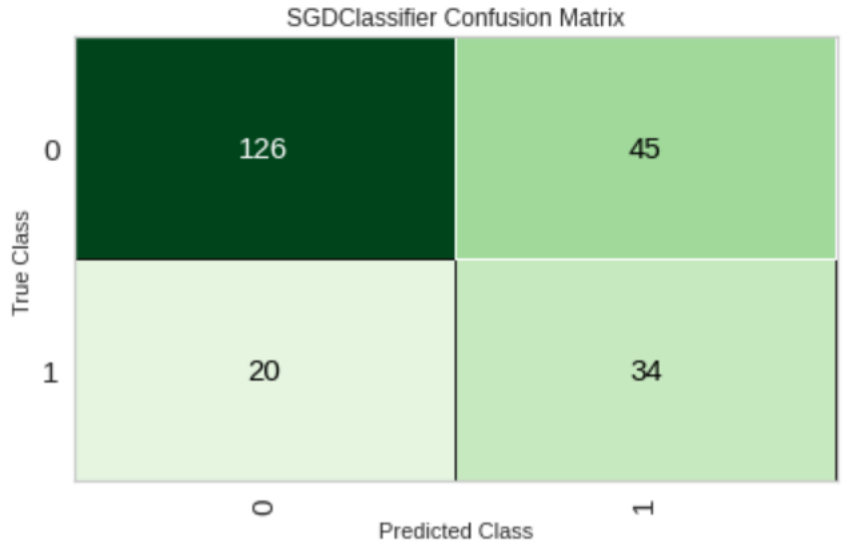
Fold	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.7547	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.7547	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.7547	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.7692	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.5577	0.0000	0.9167	0.3333	0.4889	0.2274	0.3208
5	0.7692	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.7692	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.7692	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.3462	0.0000	1.0000	0.2609	0.4138	0.0753	0.1978
9	0.4615	0.0000	1.0000	0.3171	0.4815	0.1642	0.2990
Mean	0.6706	0.0000	0.2917	0.0911	0.1384	0.0467	0.0818
Std	0.1489	0.0000	0.4460	0.1402	0.2122	0.0791	0.1283

```
[ ] pip install matplotlib==3.0.3
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: matplotlib==3.0.3 in /usr/local/lib/python3.7/dist-packages (3.0.3)
Requirement already satisfied: numpy>=1.10.0 in /usr/local/lib/python3.7/dist-packages (from matplotlib==3.0.3) (1.21.6)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages (from matplotlib==3.0.3) (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==3.0.3) (3.0.9)
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib==3.0.3) (2.8.2)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib==3.0.3) (1.4.4)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.7/dist-packages (from kiwisolver>=1.0.1->matplotlib==3.0.3) (4.1.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (from python-dateutil>=2.1->matplotlib==3.0.3) (1.15.0)
```

```
[ ] from google.colab import drive
    drive.mount('/content/drive')
```

```
[ ]  plot_model(sModel, plot='confusion_matrix')
```



GIT HUB LINK - https://github.com/Drishti412/Project/upload

Google drive link – https://drive.google.com/drive/folders/1l4dMaYvw6iqDz5SPeXB9KHvEJE-lZL-1?usp=sharing