Model

```
!pip install pycaret
```

```
Requirement already satisfied: orjson<4.0.0,>=3.8.0 in /usr/local/lib/python3.10/dist-packages (from plotly-resampler>=0.8.3.1->p
Requirement already satisfied: tsdownsample>=0.1.3 in /usr/local/lib/python3.10/dist-packages (from plotly-resampler>=0.8.3.1->py
Requirement already satisfied: Cython!=0.29.18,!=0.29.31,>=0.29 in /usr/local/lib/python3.10/dist-packages (from pmdarima>=2.0.4-
Requirement already satisfied: urllib3 in /usr/local/lib/python3.10/dist-packages (from pmdarima>=2.0.4->pycaret) (2.2.3)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests>=2.27.1->pycare
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests>=2.27.1->pycaret) (3.10)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests>=2.27.1->pycaret) (20
Requirement already satisfied: Flask<3.1,>=1.0.4 in /usr/local/lib/python3.10/dist-packages (from dash>=2.9.0->plotly-resampler>=
Requirement already satisfied: Werkzeug<3.1 in /usr/local/lib/python3.10/dist-packages (from dash>=2.9.0->plotly-resampler>=0.8.3
Requirement already satisfied: dash-html-components==2.0.0 in /usr/local/lib/python3.10/dist-packages (from dash>=2.9.0->plotly-r
Requirement already satisfied: dash-core-components==2.0.0 in /usr/local/lib/python3.10/dist-packages (from dash>=2.9.0->plotly-r
Requirement already satisfied: dash-table==5.0.0 in /usr/local/lib/python3.10/dist-packages (from dash>=2.9.0->plotly-resampler>=
Requirement already satisfied: typing-extensions>=4.1.1 in /usr/local/lib/python3.10/dist-packages (from dash>=2.9.0->plotly-resa
Requirement already satisfied: retrying in /usr/local/lib/python3.10/dist-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->
Requirement already satisfied: nest-asyncio in /usr/local/lib/python3.10/dist-packages (from dash>=2.9.0->plotly-resampler>=0.8.3
Requirement already satisfied: jupyter-client in /usr/local/lib/python3.10/dist-packages (from ipykernel>=4.5.1->ipywidgets>=7.6.
Requirement already satisfied: tornado>=4.2 in /usr/local/lib/python3.10/dist-packages (from ipykernel>=4.5.1->ipywidgets>=7.6.5-
Requirement already satisfied: parso<0.9.0,>=0.8.3 in /usr/local/lib/python3.10/dist-packages (from jedi>=0.16->ipython>=5.5.0->p
Requirement already satisfied: attrs>=22.2.0 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat>=4.2.0->p
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=
Requirement already satisfied: referencing>=0.28.4 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat>=4.
Requirement already satisfied: rpds-py>=0.7.1 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat>=4.2.0->
Requirement already satisfied: platformdirs>=2.5 in /usr/local/lib/python3.10/dist-packages (from jupyter-core!=5.0.*,>=4.12->nbf
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages (from patsy>=0.5.1->category-encoders>=2.4.0->pycar
Requirement already satisfied: ptyprocess>=0.5 in /usr/local/lib/python3.10/dist-packages (from pexpect>4.3->ipython>=5.5.0->pyca
Requirement already satisfied: wcwidth in /usr/local/lib/python3.10/dist-packages (from prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.
Requirement already satisfied: notebook>=4.4.1 in /usr/local/lib/python3.10/dist-packages (from widgetsnbextension~=3.6.0->ipywid
Requirement already satisfied: itsdangerous>=2.0 in /usr/local/lib/python3.10/dist-packages (from Flask<3.1,>=1.0.4->dash>=2.9.0-
Requirement already satisfied: click>=8.0 in /usr/local/lib/python3.10/dist-packages (from Flask<3.1,>=1.0.4->dash>=2.9.0->plotly
Requirement already satisfied: pyzmq<25,>=17 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension
Requirement already satisfied: argon2-cffi in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=
Requirement already satisfied: nbconvert>=5 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~
Requirement already satisfied: Send2Trash>=1.8.0 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbexten
Requirement already satisfied: terminado>=0.8.3 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextens
Requirement already satisfied: prometheus-client in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbexten
Requirement already satisfied: nbclassic>=0.4.7 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextens
Requirement already satisfied: notebook-shim>=0.2.3 in /usr/local/lib/python3.10/dist-packages (from nbclassic>=0.4.7->notebook>=
Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbexte
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->wid
Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbex
Requirement already satisfied: defusedxml in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgets
Requirement already satisfied: entrypoints>=0.2.2 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1-
Requirement already satisfied: jupyterlab-pygments in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1
Requirement already satisfied: mistune<2,>=0.8.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->
Requirement already satisfied: nbclient>=0.5.0 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->wi
Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.
Requirement already satisfied: tinycss2 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnb
Requirement already satisfied: argon2-cffi-bindings in /usr/local/lib/python3.10/dist-packages (from argon2-cffi->notebook>=4.4.1
Requirement already satisfied: jupyter-server<3,>=1.8 in /usr/local/lib/python3.10/dist-packages (from notebook-shim>=0.2.3->nbcl
Requirement already satisfied: cffi>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from argon2-cffi-bindings->argon2-cffi->no
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-packages (from beautifulsoup4->nbconvert>=5->noteb
Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-packages (from bleach->nbconvert>=5->notebook>=4.4.
Requirement already satisfied: pycparser in /usr/local/lib/python3.10/dist-packages (from cffi>=1.0.1->argon2-cffi-bindings->argo
Requirement already satisfied: anyio<4,>=3.1.0 in /usr/local/lib/python3.10/dist-packages (from jupyter-server<3,>=1.8->notebook-
Requirement already satisfied: websocket-client in /usr/local/lib/python3.10/dist-packages (from jupyter-server<3,>=1.8->notebook
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.10/dist-packages (from anyio<4,>=3.1.0->jupyter-server<3,>=
Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-packages (from anyio<4,>=3.1.0->jupyter-server<3,
```

```
from pycaret.classification import *
```

```
import pandas as pd

df = pd.read_csv('https://raw.githubusercontent.com/enzoschitini/Git-Large-File-Storage/refs/heads/main/olist_model_data.csv').drop(colu
df.head()
```

	customer_state	product_id	product_category_name	product_name_lenght	product_description_lenght	produc
0	RJ	6490b9f63b0b595e1c1402b46b0f50ad	telefonia	24.0	820.0	
1	MG	1dec4c88c685d5a07bf01dcb0f8bf9f8	automotivo	44.0	982.0	
2	PE	c0315bd17c58be3dec2bd04dd4233b7e	perfumaria	59.0	827.0	
3	SP	163da86e86f0089de089399c28c3fcbb	informatica_acessorios	60.0	407.0	
4	ES	ee406bf28024d97771c4b1e8b7e8e219	cool_stuff	54.0	629.0	

```
df.shape
```

```
(28182, 18)
```

```
df = df.select_dtypes('number')


import numpy as np

# Replace infinite values with NaN
df.replace([np.inf, -np.inf], np.nan, inplace=True)

# Option 1: Remove rows with NaN values
df.dropna(inplace=True)

# Option 2: Alternatively, fill NaN values with the mean or median (for numerical columns)
df.fillna(df.mean(), inplace=True)

# After cleaning, run the setup again
clf = setup(df, target='default', session_id=123,
            preprocess=True,
            remove_outliers=True,
            normalize=True,
            feature_selection=True)
```

```
[LightGBM] [Info] Number of positive: 9299, number of negative: 9440
[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing was 0.000762 seconds.
You can set `force_row_wise=true` to remove the overhead.
And if memory is not enough, you can set `force_col_wise=true`.
[LightGBM] [Info] Total Bins 1843
[LightGBM] [Info] Number of data points in the train set: 18739, number of used features: 12
[LightGBM] [Info] [binary:BoostFromScore]: pavg=0.496238 -> initscore=-0.015049
[LightGBM] [Info] Start training from score -0.015049
```

	Description	Value
0	Session id	123
1	Target	default
2	Target type	Binary
3	Original data shape	(28181, 13)
4	Transformed data shape	(27194, 3)
5	Transformed train set shape	(18739, 3)
6	Transformed test set shape	(8455, 3)
7	Numeric features	12
8	Preprocess	True
9	Imputation type	simple
10	Numeric imputation	mean
11	Categorical imputation	mode
12	Remove outliers	True
13	Outliers threshold	0.050000
14	Normalize	True
15	Normalize method	zscore
16	Feature selection	True
17	Feature selection method	classic
18	Feature selection estimator	lightgbm
19	Number of features selected	0.200000
20	Fold Generator	StratifiedKFold
21	Fold Number	10
22	CPU Jobs	-1
23	Use GPU	False
24	Log Experiment	False
25	Experiment Name	clf-default-name
26	USI	2891

```
best_model = compare_models()
```

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
lightgbm	Light Gradient Boosting Machine	0.9209	0.9490	0.9482	0.8991	0.9230	0.8417	0.8430	0.9580
xgboost	Extreme Gradient Boosting	0.9195	0.9446	0.9518	0.8940	0.9220	0.8390	0.8408	0.8480
knn	K Neighbors Classifier	0.9177	0.9342	0.9534	0.8900	0.9206	0.8354	0.8376	0.6950
gbc	Gradient Boosting Classifier	0.9167	0.9475	0.9305	0.9055	0.9178	0.8334	0.8338	1.7020
svm	SVM - Linear Kernel	0.9116	0.9314	0.9178	0.9067	0.9120	0.8232	0.8236	0.7590
lr	Logistic Regression	0.9094	0.9320	0.9111	0.9080	0.9095	0.8187	0.8188	1.1550
rf	Random Forest Classifier	0.8997	0.9372	0.9137	0.8889	0.9011	0.7995	0.7998	2.4250
et	Extra Trees Classifier	0.8864	0.9123	0.8887	0.8847	0.8866	0.7728	0.7728	1.4960
lda	Linear Discriminant Analysis	0.8831	0.9303	0.8485	0.9117	0.8789	0.7663	0.7683	0.5140
ridge	Ridge Classifier	0.8831	0.9303	0.8485	0.9117	0.8789	0.7663	0.7683	0.5300
ada	Ada Boost Classifier	0.8820	0.9223	0.8783	0.8849	0.8815	0.7640	0.7641	1.0900
dt	Decision Tree Classifier	0.8596	0.8812	0.8483	0.8680	0.8580	0.7193	0.7195	0.5460
qda	Quadratic Discriminant Analysis	0.8042	0.9353	0.6640	0.9227	0.7721	0.6084	0.6339	0.5250
nb	Naive Bayes	0.7233	0.8413	0.4850	0.9263	0.6362	0.4465	0.5078	0.6520
dummy	Dummy Classifier	0.5000	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.6980

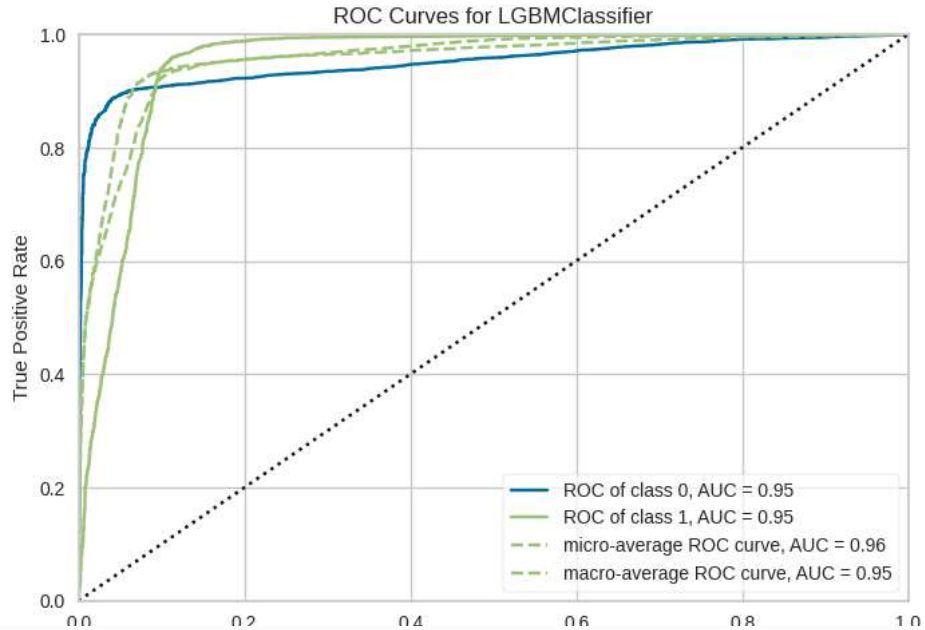
```
model = create_model('lightgbm')
```

Fold	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.9265	0.9501	0.9625	0.8979	0.9291	0.8530	0.8552
1	0.9255	0.9538	0.9534	0.9031	0.9276	0.8510	0.8523
2	0.9235	0.9518	0.9463	0.9050	0.9252	0.8469	0.8478
3	0.9179	0.9495	0.9432	0.8977	0.9199	0.8358	0.8369
4	0.9154	0.9410	0.9452	0.8919	0.9178	0.8307	0.8322
5	0.9179	0.9485	0.9442	0.8969	0.9200	0.8358	0.8370
6	0.9214	0.9528	0.9483	0.8999	0.9235	0.8428	0.8440
7	0.9123	0.9340	0.9412	0.8897	0.9147	0.8245	0.8259
8	0.9214	0.9493	0.9483	0.8999	0.9235	0.8428	0.8440
9	0.9270	0.9592	0.9493	0.9087	0.9286	0.8540	0.8548
Mean	0.9209	0.9490	0.9482	0.8991	0.9230	0.8417	0.8430
Std	0.0047	0.0066	0.0058	0.0054	0.0046	0.0094	0.0094

```
evaluate_model(model)
```

Plot Type:

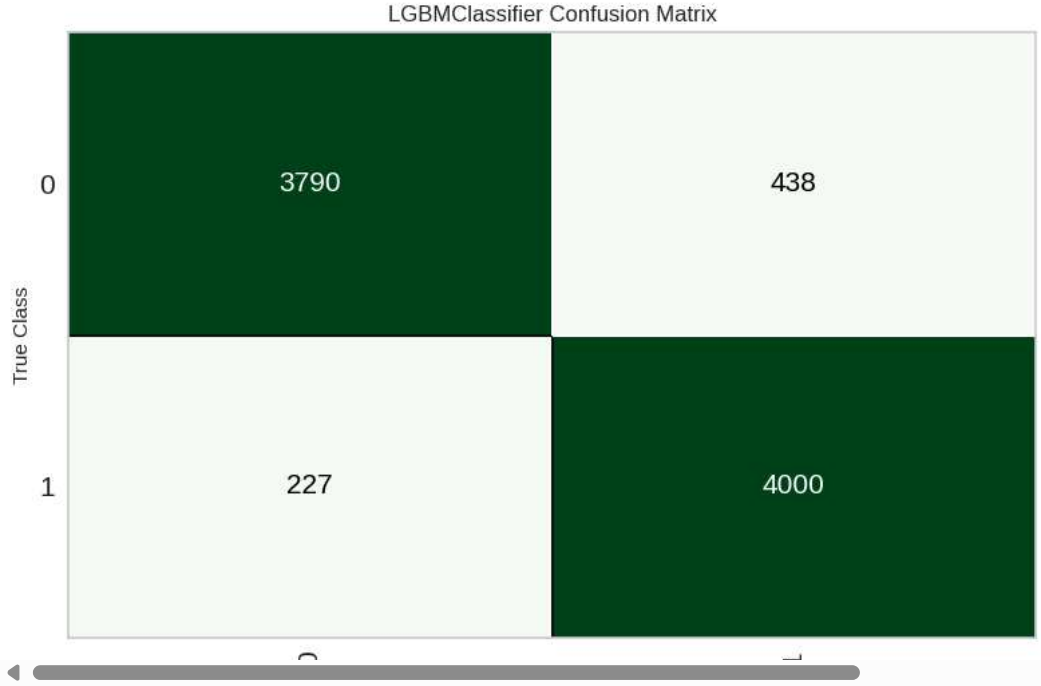
Pipeline Plot	Hyperparameters	AUC	Confusion Matrix	Threshold	Precision Recall
Prediction Error	Class Report	Feature Selection	Learning Curve	Manifold Learning	Calibration Curve
Validation Curve	Dimensions	Feature Importance	Feature Importance…	Decision Boundary	Lift Chart
Gain Chart	Decision Tree	KS Statistic Plot			




```
evaluate_model(model)
```

Plot Type:

Pipeline Plot	Hyperparameters	AUC	Confusion Matrix	Threshold	Precision Recall
Prediction Error	Class Report	Feature Selection	Learning Curve	Manifold Learning	Calibration Curve
Validation Curve	Dimensions	Feature Importance	Feature Importance…	Decision Boundary	Lift Chart
Gain Chart	Decision Tree	KS Statistic Plot			




```
evaluate_model(model)
```

Plot Type:

Pipeline Plot	Hyperparameters	AUC	Confusion Matrix	Threshold	Precision Recall
Prediction Error	Class Report	Feature Selection	Learning Curve	Manifold Learning	Calibration Curve
Validation Curve	Dimensions	Feature Importance	Feature Importance…	Decision Boundary	Lift Chart
Gain Chart	Decision Tree	KS Statistic Plot			

```
evaluate_model(model)
```

Plot Type:

Pipeline Plot	Hyperparameters	AUC	Confusion Matrix	Threshold	Precision Recall
Prediction Error	Class Report	Feature Selection	Learning Curve	Manifold Learning	Calibration Curve
Validation Curve	Dimensions	Feature Importance	Feature Importance…	Decision Boundary	Lift Chart
Gain Chart	Decision Tree	KS Statistic Plot			

Parameters	
boosting_type	gbdt
class_weight	None
colsample_bytree	1.0
importance_type	split
learning_rate	0.1
max_depth	-1
min_child_samples	20
min_child_weight	0.001
min_split_gain	0.0
n_estimators	100
n_jobs	-1
num_leaves	31
objective	None
random_state	123
reg_alpha	0.0
reg_lambda	0.0