In [2]:

```python
import numpy as np
import pandas as pd
import os
```

In [4]:

```python
train=pd.read_csv("train.csv")
```

In [7]:

```python
train.head()
```

Out[7]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	

In [8]:

```python
test=pd.read_csv("test.csv")
```

In [9]:

```python
import pycaret
```

PassengerID is of no such use hence will be ignored by lib

Name and ticket will also be ignored while modeling

In [13]:

```python
from pycaret.classification import *
clf1=setup(train,target="Survived",ignore_features=['Name','PassengerId','Ticket'])
```

Setup Succesfully Completed!

	Description	Value
0	session_id	796
1	Target Type	Binary
2	Label Encoded	None
3	Original Data	(891, 12)
4	Missing Values	True
5	Numeric Features	3
6	Categorical Features	8
7	Ordinal Features	False
8	High Cardinality Features	False
9	High Cardinality Method	None
10	Sampled Data	(891, 12)
11	Transformed Train Set	(623, 171)
12	Transformed Test Set	(268, 171)
13	Numeric Imputer	mean
14	Categorical Imputer	constant
15	Normalize	False
16	Normalize Method	None
17	Transformation	False
18	Transformation Method	None
19	PCA	False
20	PCA Method	None
21	PCA Components	None
22	Ignore Low Variance	False
23	Combine Rare Levels	False
24	Rare Level Threshold	None
25	Numeric Binning	False
26	Remove Outliers	False
27	Outliers Threshold	None
28	Remove Multicollinearity	False
29	Multicollinearity Threshold	None
30	Clustering	False
31	Clustering Iteration	None
32	Polynomial Features	False

	Description	Value
33	Polynomial Degree	None
34	Trignometry Features	False
35	Polynomial Threshold	None
36	Group Features	False
37	Feature Selection	False
38	Features Selection Threshold	None
39	Feature Interaction	False
40	Feature Ratio	False
41	Interaction Threshold	None

Comparing Different classification models

In [14]:

```
compare_models()
```

Out[14]:

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa
0	Gradient Boosting Classifier	0.820200	0.859000	0.666100	0.840900	0.738000	0.604700
1	CatBoost Classifier	0.818600	0.870000	0.661800	0.841500	0.736100	0.601300
2	Extreme Gradient Boosting	0.812200	0.857800	0.753400	0.766900	0.755300	0.603600
3	Logistic Regression	0.805700	0.850600	0.703600	0.776700	0.734300	0.582200
4	Light Gradient Boosting Machine	0.804100	0.868800	0.711800	0.770400	0.734400	0.580500
5	Ridge Classifier	0.797700	0.000000	0.682800	0.767200	0.720900	0.563400
6	Linear Discriminant Analysis	0.786500	0.837200	0.670300	0.751700	0.706000	0.539600
7	Extra Trees Classifier	0.784800	0.837800	0.703600	0.740700	0.712400	0.542200
8	Random Forest Classifier	0.780100	0.839600	0.661800	0.750800	0.697200	0.526300
9	Ada Boost Classifier	0.772000	0.832200	0.707600	0.706300	0.704800	0.519500
10	Decision Tree Classifier	0.770500	0.764700	0.728300	0.700500	0.708900	0.520800
11	K Neighbors Classifier	0.719300	0.736100	0.569200	0.665400	0.610900	0.393000
12	SVM - Linear Kernel	0.679200	0.000000	0.678400	0.646700	0.601800	0.346900
13	Naive Bayes	0.646900	0.769900	0.121400	0.738100	0.206600	0.112800
14	Quadratic Discriminant Analysis	0.621400	0.595500	0.518100	0.571300	0.499600	0.214500

GD classifier gives highest accuracy

In [15]:

```
tuned_lightgbm = tune_model('lightgbm',optimize='AUC')
```

	Accuracy	AUC	Recall	Prec.	F1	Kappa
0	0.7937	0.8600	0.6250	0.7895	0.6977	0.5442
1	0.7937	0.8269	0.7500	0.7200	0.7347	0.5660
2	0.8413	0.8349	0.6667	0.8889	0.7619	0.6465
3	0.8548	0.8586	0.7500	0.8571	0.8000	0.6869
4	0.7903	0.8372	0.6250	0.7895	0.6977	0.5405
5	0.7742	0.8010	0.5417	0.8125	0.6500	0.4930
6	0.8226	0.8860	0.7083	0.8095	0.7556	0.6173
7	0.7742	0.8311	0.7083	0.7083	0.7083	0.5241
8	0.8548	0.9474	0.8750	0.7778	0.8235	0.7010
9	0.8226	0.8941	0.7826	0.7500	0.7660	0.6232
Mean	0.8122	0.8577	0.7033	0.7903	0.7395	0.5943
SD	0.0296	0.0400	0.0894	0.0534	0.0497	0.0674

In [16]:

```
evaluate_model(tuned_lightgbm)
```

A Jupyter widget could not be displayed because the widget state could not be found. This could happen if the kernel storing the widget is no longer available, or if the widget state was not saved in the notebook. You may be able to create the widget by running the appropriate cells.

In [17]:

```
final_lightgbm=finalize_model(tuned_lightgbm)
```

In [18]:

```
print(final_lightgbm)
```

```
LGBMClassifier(boosting_type='gbdt', class_weight=None, colsample_bytree=1.
0,
               importance_type='split', learning_rate=0.6, max_depth=90,
               min_child_samples=20, min_child_weight=0.001, min_split_gain=
0.2,
               n_estimators=30, n_jobs=-1, num_leaves=150, objective=None,
               random_state=796, reg_alpha=0.4, reg_lambda=0.2, silent=True,
               subsample=1.0, subsample_for_bin=200000, subsample_freq=0)
```

In [19]:

```
predictions = predict_model(final_lightgbm, data=test)
```

In [20]:

```
predictions.head()
```

Out[20]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Emba
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	

In [21]:

```
predictions
```

Out[21]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabir
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN
...	...	...	...	...	...	...	...	...	...	..
413	1305	3	Spector, Mr. Woolf	male	NaN	0	0	A.5. 3236	8.0500	NaN
414	1306	1	Oliva y Ocana, Dona. Fermina	female	39.0	0	0	PC 17758	108.9000	C105
415	1307	3	Saether, Mr. Simon Sivertsen	male	38.5	0	0	SOTON/O.Q. 3101262	7.2500	NaN
416	1308	3	Ware, Mr. Frederick	male	NaN	0	0	359309	8.0500	NaN
417	1309	3	Peter, Master. Michael J	male	NaN	1	1	2668	22.3583	NaN

418 rows × 13 columns

In [ ]: