Predictive Modeling for Chronic obstructive pulmonary disease Outcomes using Machine Learning

Problem

Chronic Obstructive Pulmonary Disease (COPD) is a progressive respiratory condition characterized by persistent airflow limitation, significantly affecting millions worldwide. Despite being preventable and manageable, COPD often remains underdiagnosed, especially in its early stages, due to its subtle symptoms and overlapping risk factors such as smoking, environmental pollutants, and genetic factors. Traditional diagnostic approaches, such as spirometry, are often conducted late in the disease progression, limiting opportunities for early intervention. Furthermore, the complex nature of COPD makes its prediction challenging, often leading to delays in appropriate care. There is a critical need for innovative solutions that leverage machine learning to analyze diverse patient data, predict disease onset accurately, and identify individuals at high risk. Developing such models can revolutionize COPD management by facilitating timely interventions, personalized treatment plans, and ultimately improving patient outcomes.

```
!pip install researchpy
```

```
Requirement already satisfied: researchpy in /usr/local/lib/python3.10/dist-packages (0.3.6)
Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages (from researchpy) (1.11.4)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from researchpy) (1.26.4)
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from researchpy) (2.2.2)
Requirement already satisfied: statsmodels in /usr/local/lib/python3.10/dist-packages (from researchpy) (0.14.1)
Requirement already satisfied: patsy in /usr/local/lib/python3.10/dist-packages (from researchpy) (0.5.6)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.10/dist-packages (from pandas->researchpy) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas->researchpy) (2024.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.10/dist-packages (from pandas->researchpy) (2024.1)
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages (from patsy->researchpy) (1.16.0)
Requirement already satisfied: packaging>=21.3 in /usr/local/lib/python3.10/dist-packages (from statsmodels->researchpy) (24.1)
```

```
# Load libraries
import numpy as np
import pandas as pd
import researchpy as rp
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Load dataset
data = pd.read_csv("/content/COPD dataset.csv")
data.head()
```

	SL	ID	AGE	PackHistory	COPDSEVERITY	MWT1	MWT2	MWTBest	FEV1	FEV1PRED	...	SGRQ	AGEquartiles	copd	gender	smoking	Diabetes	muscular	hypertension	AtrialFib	Outcome
0	1	58	77	60.0	SEVERE	120.0	120.0	120.0	1.21	36.0	...	69.55	4	3	1	2	1	0	0	1	0
1	2	57	79	50.0	MODERATE	165.0	176.0	176.0	1.09	56.0	...	44.24	4	2	0	2	1	0	0	1	1
2	3	62	80	11.0	MODERATE	201.0	180.0	201.0	1.52	68.0	...	44.09	4	2	0	2	1	0	0	1	0
3	4	145	56	60.0	VERY SEVERE	210.0	210.0	210.0	0.47	14.0	...	62.04	1	4	1	2	0	0	1	1	0
4	5	136	65	68.0	SEVERE	204.0	210.0	210.0	1.07	42.0	...	75.56	1	3	1	2	0	1	1	0	0

5 rows × 24 columns

```
data.drop('SL', axis=1, inplace=True)
```

```
data.drop('ID', axis=1, inplace=True)
```

```
data.head()
```

	AGE	PackHistory	COPDSEVERITY	MWT1	MWT2	MWTBest	FEV1	FEV1PRED	FVC	FVCPRED	...	SGRQ	AGEquartiles	copd	gender	smoking	Diabetes	muscular	hypertension	AtrialFib	Outcome
0	77	60.0	SEVERE	120.0	120.0	120.0	1.21	36.0	2.40	98	...	69.55	4	3	1	2	1	0	0	1	0
1	79	50.0	MODERATE	165.0	176.0	176.0	1.09	56.0	1.64	65	...	44.24	4	2	0	2	1	0	0	1	1
2	80	11.0	MODERATE	201.0	180.0	201.0	1.52	68.0	2.30	86	...	44.09	4	2	0	2	1	0	0	1	0
3	56	60.0	VERY SEVERE	210.0	210.0	210.0	0.47	14.0	1.14	27	...	62.04	1	4	1	2	0	0	1	1	0
4	65	68.0	SEVERE	204.0	210.0	210.0	1.07	42.0	2.91	98	...	75.56	1	3	1	2	0	1	1	0	0

5 rows × 22 columns

Exploring Data

```
# check shape of data
data.shape
```

```
(101, 22)
```

```
# dtypes
data.dtypes
```

	0
AGE	int64
PackHistory	float64
COPDSEVERITY	object
MWT1	float64
MWT2	float64
MWTBest	float64
FEV1	float64
FEV1PRED	float64
FVC	float64
FVCPRED	int64
CAT	int64
HAD	float64
SGRQ	float64
AGEquartiles	int64
copd	int64
gender	int64
smoking	int64
Diabetes	int64
muscular	int64
hypertension	int64
AtrialFib	int64
Outcome	int64

dtype: object

```
# info
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 101 entries, 0 to 100
Data columns (total 22 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   AGE          101 non-null    int64
 1   PackHistory  101 non-null    float64
 2   COPDSEVERITY 101 non-null    object
 3   MWT1         99 non-null     float64
 4   MWT2         100 non-null    float64
 5   MWTBest      101 non-null    float64
 6   FEV1         101 non-null    float64
 7   FEV1PRED     101 non-null    float64
 8   FVC          101 non-null    float64
 9   FVCPRED      101 non-null    int64
 10  CAT          101 non-null    int64
 11  HAD          101 non-null    float64
 12  SGRQ         101 non-null    float64
 13  AGEquartiles 101 non-null    int64
 14  copd         101 non-null    int64
 15  gender       101 non-null    int64
 16  smoking      101 non-null    int64
 17  Diabetes     101 non-null    int64
 18  muscular     101 non-null    int64
 19  hypertension 101 non-null    int64
 20  AtrialFib    101 non-null    int64
 21  Outcome      101 non-null    int64
dtypes: float64(9), int64(12), object(1)
memory usage: 17.5+ KB
```

```
# check missing data
data.isnull().sum()
```

	0
AGE	0
PackHistory	0
COPDSEVERITY	0
MWT1	2
MWT2	1
MWTBest	1
FEV1	0
FEV1PRED	0
FVC	0
FVCPRED	0
CAT	0
HAD	0
SGRQ	0
AGEquartiles	0
copd	0
gender	0
smoking	0
Diabetes	0
muscular	0
hypertension	0
AtrialFib	0
Outcome	0

dtype: int64

```
data = data.dropna()
```

```
data[['COPDSEVERITY','AGEquartiles', 'copd', 'gender', 'smoking', 'Diabetes', 'muscular','hypertension', 'AtrialFib', 'Outcome']] = data[['COPDSEVERITY','AGEquarti
```

```
<ipython-input-12-670428916c6a>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  data[['COPDSEVERITY','AGEquartiles', 'copd', 'gender', 'smoking', 'Diabetes', 'muscular','hypertension', 'AtrialFib', 'Outcome']] = data[['COPDSEVERITY','AGEquartiles', 'copd', 'gender', 'smoking', 'Diabetes', 'muscular','hypertension', 'AtrialFib', 'Outcome']].astype(object)
```

```
data.isnull().sum()
```

	0
AGE	0
PackHistory	0
COPDSEVERITY	0
MWT1	0
MWT2	0
MWTBest	0
FEV1	0
FEV1PRED	0
FVC	0
FVCPRED	0
CAT	0
HAD	0
SGRQ	0
AGEquartiles	0
copd	0
gender	0
smoking	0
Diabetes	0
muscular	0
hypertension	0
AtrialFib	0
Outcome	0

dtype: int64

Descriptive Statistic

```
# select numeric data
num_cols = data.select_dtypes(exclude = 'object')
num_cols.head()
```

	AGE	PackHistory	MWT1	MWT2	MWTBest	FEV1	FEV1PRED	FVC	FVCPRED	CAT	HAD	SGRQ
0	77	60.0	120.0	120.0	120.0	1.21	36.0	2.40	98	25	8.0	69.55
1	79	50.0	165.0	176.0	176.0	1.09	56.0	1.64	65	12	21.0	44.24
2	80	11.0	201.0	180.0	201.0	1.52	68.0	2.30	86	22	18.0	44.09
3	56	60.0	210.0	210.0	210.0	0.47	14.0	1.14	27	28	26.0	62.04
4	65	68.0	204.0	210.0	210.0	1.07	42.0	2.91	98	32	18.0	75.56

Next steps: Generate code with num_cols View recommended plots New interactive sheet

```
num_cols.columns
```

```
Index(['AGE', 'PackHistory', 'MWT1', 'MWT2', 'MWTBest', 'FEV1', 'FEV1PRED',
       'FVC', 'FVCPRED', 'CAT', 'HAD', 'SGRQ'],
      dtype='object')
```

```
# summary statistics of numerical variables
rp.summarize(num_cols[['AGE', 'PackHistory', 'MWT1', 'MWT2', 'MWTBest', 'FEV1', 'FEV1PRED',
       'FVC', 'FVCPRED', 'CAT', 'HAD', 'SGRQ']])
```

	Name	N	Mean	Median	Variance	SD	SE	95% Conf. Interval
0	AGE	99	70.1212	71.0	61.9852	7.8731	0.7913	AGE
1	PackHistory	99	39.803	36.0	606.3001	24.6232	2.4747	PackHistory
2	MWT1	99	385.8586	419.0	10971.3471	104.7442	10.5272	MWT1
3	MWT2	99	389.798	399.0	11706.0404	108.1945	10.874	MWT2
4	MWTBest	99	398.7172	420.0	11453.1845	107.0196	10.7559	MWTBest
5	FEV1	99	1.6031	1.6	0.457	0.676	0.0679	FEV1
6	FEV1PRED	99	58.4008	60.0	503.8626	22.4469	2.256	FEV1PRED
7	FVC	99	2.9578	2.77	0.9589	0.9793	0.0984	FVC
8	FVCPRED	99	86.3434	84.0	481.1053	21.9341	2.2045	FVCPRED
9	CAT	99	19.404	18.0	355.5494	18.856	1.8951	CAT
10	HAD	99	11.1131	10.0	73.8956	8.5963	0.864	HAD
11	SGRQ	99	40.1696	38.21	337.075	18.3596	1.8452	SGRQ

```
# select categorical data
cat_cols = data.select_dtypes(include = 'object')
cat_cols.head()
```

	COPDSEVERITY	AGEquartiles	copd	gender	smoking	Diabetes	muscular	hypertension	AtrialFib	Outcome
0	SEVERE	4	3	1	2	1	0	0	1	0
1	MODERATE	4	2	0	2	1	0	0	1	1
2	MODERATE	4	2	0	2	1	0	0	1	0
3	VERY SEVERE	1	4	1	2	0	0	1	1	0
4	SEVERE	1	3	1	2	0	1	1	0	0

Next steps: Generate code with cat_cols View recommended plots New interactive sheet

```
cat_cols.columns
```

```
Index(['COPDSEVERITY', 'AGEquartiles', 'copd', 'gender', 'smoking', 'Diabetes',
       'muscular', 'hypertension', 'AtrialFib', 'Outcome'],
      dtype='object')
```

```
# summary statistics of categorical variables
rp.summary_cat(cat_cols[['COPDSEVERITY', 'AGEquartiles', 'copd', 'gender', 'smoking', 'Diabetes',
       'muscular', 'hypertension', 'AtrialFib', 'Outcome']])
```

	Variable	Outcome	Count	Percent
0	COPDSEVERITY	MODERATE	41	41.41
1		SEVERE	27	27.27
2		MILD	23	23.23
3		VERY SEVERE	8	8.08
4	AGEquartiles	3	28	28.28
5		1	25	25.25
6		2	24	24.24
7		4	22	22.22
8	copd	2	41	41.41
9		3	27	27.27
10		1	23	23.23
11		4	8	8.08
12	gender	1	63	63.64
13		0	36	36.36
14	smoking	2	83	83.84
15		1	16	16.16
16	Diabetes	0	78	78.79
17		1	21	21.21
18	muscular	0	80	80.81
19		1	19	19.19
20	hypertension	0	88	88.89
21		1	11	11.11
22	AtrialFib	0	79	79.80
23		1	20	20.20
24	Outcome	0	90	90.91
25		1	9	9.09

Correlations between Variable

```
# correlation: Pearson's by default
num_cols.corr(method='pearson')
```

	AGE	PackHistory	MWT1	MWT2	MWTBest	FEV1	FEV1PRED	FVC	FVCPRED	CAT	HAD	SGRQ
AGE	1.000000	-0.021614	-0.246847	-0.234462	-0.227397	-0.087169	0.080862	-0.129459	0.014883	0.081531	-0.210833	-0.127909
PackHistory	-0.021614	1.000000	-0.251895	-0.280635	-0.249855	-0.120797	-0.123172	-0.081097	0.001033	-0.146967	0.044868	0.042651
MWT1	-0.246847	-0.251895	1.000000	0.954434	0.982371	0.458655	0.364297	0.449893	0.257159	-0.135281	-0.262240	-0.507541
MWT2	-0.234462	-0.280635	0.954434	1.000000	0.982244	0.470493	0.408255	0.456276	0.299864	-0.153157	-0.293162	-0.522844
MWTBest	-0.227397	-0.249855	0.982371	0.982244	1.000000	0.466083	0.388947	0.444527	0.258128	-0.153593	-0.294638	-0.537424
FEV1	-0.087169	-0.120797	0.458655	0.470493	0.466083	1.000000	0.773323	0.818421	0.514994	-0.063451	-0.163453	-0.314719
FEV1PRED	0.080862	-0.123172	0.364297	0.408255	0.388947	0.773323	1.000000	0.519598	0.624727	0.009657	-0.121885	-0.340204
FVC	-0.129459	-0.081097	0.449893	0.456276	0.444527	0.818421	0.519598	1.000000	0.623630	-0.158801	-0.145300	-0.232152
FVCPRED	0.014883	0.001033	0.257159	0.299864	0.258128	0.514994	0.624727	0.623630	1.000000	-0.137885	-0.160940	-0.294018
CAT	0.081531	-0.146967	-0.135281	-0.153157	-0.153593	-0.063451	0.009657	-0.158801	-0.137885	1.000000	0.166804	0.290358
HAD	-0.210833	0.044868	-0.262240	-0.293162	-0.294638	-0.163453	-0.121885	-0.145300	-0.160940	0.166804	1.000000	0.390066
SGRQ	-0.127909	0.042651	-0.507541	-0.522844	-0.537424	-0.314719	-0.340204	-0.232152	-0.294018	0.290358	0.390066	1.000000

Skewness

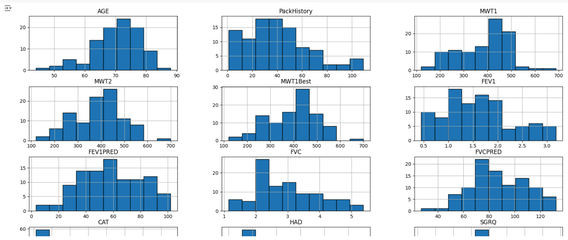
```
num_cols.skew()
```

	0
AGE	-0.737041
PackHistory	0.761152
MWT1	-0.311096
MWT2	-0.156457
MWTBest	-0.276048
FEV1	0.463046
FEV1PRED	-0.152857
FVC	0.529557
FVCPRED	0.011179
CAT	7.386027
HAD	1.807793
SGRQ	0.191143

dtype: float64

Data Visualizations

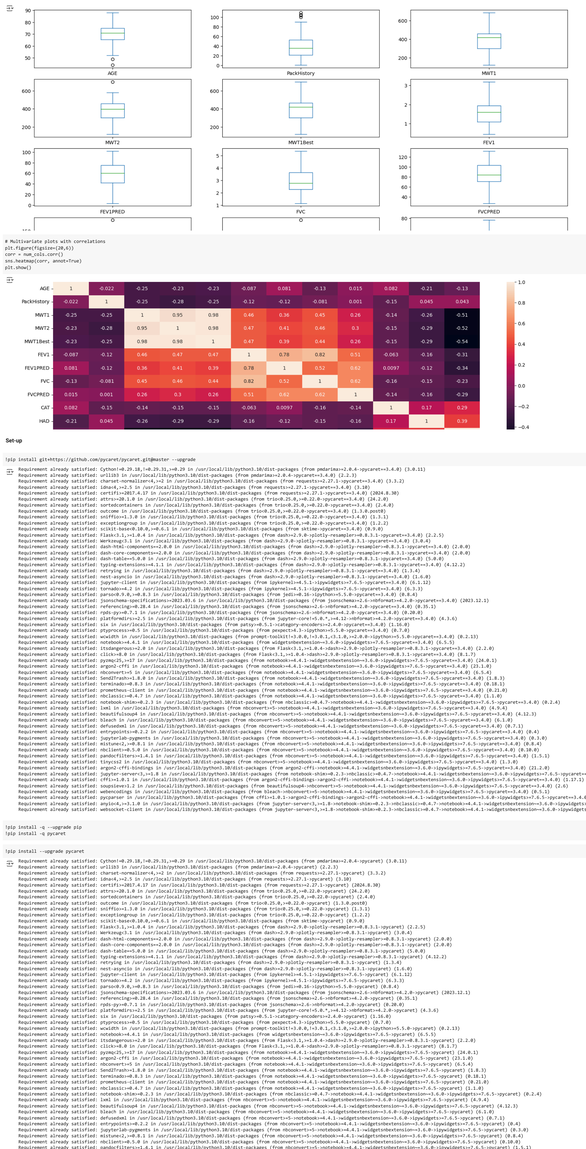
```
# Univariate distributions with histogram
data.select_dtypes(exclude = "object").hist(figsize=(20,10), edgecolor='black')
plt.show()
```




```
# Univariate distributions with density plot
data.select_dtypes(exclude = "object").plot(kind='density', subplots=True, sharex=False, figsize=(20,10), layout=(4,3))
plt.show()
```



```
# Univariate distributions with box plots
data.select_dtypes(exclude = "object").plot(kind='box', subplots=True, sharex=False, figsize=(20,10), layout=(4,3))
plt.show()
```




```
# Multivariate plots with correlations
plt.figure(figsize=(20,6))
corr = num_cols.corr()
sns.heatmap(corr, annot=True)
plt.show()
```

Set-up

```
!pip install git+https://github.com/pycaret/pycaret.git@master --upgrade
```

```
  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->pycaret==3.4.0) (3.0.11)
  Requirement already satisfied: urllib3 in /usr/local/lib/python3.10/dist-packages (from pmdarima==2.0.4->pycaret==3.4.0) (2.2.3)
  Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests!=2.27.1->pycaret==3.4.0) (3.3.2)
  Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests!=2.27.1->pycaret==3.4.0) (3.10)
  Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests!=2.27.1->pycaret==3.4.0) (2024.8.30)
  Requirement already satisfied: attrs>=20.1.0 in /usr/local/lib/python3.10/dist-packages (from trio<0.25.0,>=0.22.0->pycaret==3.4.0) (24.2.0)
  Requirement already satisfied: sortedcontainers in /usr/local/lib/python3.10/dist-packages (from trio<0.25.0,>=0.22.0->pycaret) (2.4.0)
  Requirement already satisfied: outcome in /usr/local/lib/python3.10/dist-packages (from trio<0.25.0,>=0.22.0->pycaret) (1.3.0.post0)
  Requirement already satisfied: sniffio>=1.3.0 in /usr/local/lib/python3.10/dist-packages (from trio<0.25.0,>=0.22.0->pycaret==3.4.0) (1.3.1)
  Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-packages (from trio<0.25.0,>=0.22.0->pycaret==3.4.0) (1.2.2)
  Requirement already satisfied: scikit-base<0.10.0,>=0.6.1 in /usr/local/lib/python3.10/dist-packages (from sktime->pycaret==3.4.0) (0.9.0)
  Requirement already satisfied: dash==2.9.0-splotly-resampler==0.8.3.1-cyscaret==3.4.0) (2.2.5)
  Requirement already satisfied: werkzeug==2.3.0 in /usr/local/lib/python3.10/dist-packages (from dash==2.9.0-splotly-resampler==0.8.3.1-cyscaret==3.4.0) (2.0.0)
  Requirement already satisfied: dash-core-components==2.0.0 in /usr/local/lib/python3.10/dist-packages (from dash==2.9.0-splotly-resampler==0.8.3.1-cyscaret==3.4.0) (2.0.0)
  Requirement already satisfied: dash-table==5.0.0 in /usr/local/lib/python3.10/dist-packages (from dash==2.9.0-splotly-resampler==0.8.3.1-cyscaret==3.4.0) (5.0.0)
  Requirement already satisfied: dash-html-components==2.0.0 in /usr/local/lib/python3.10/dist-packages (from dash==2.9.0-splotly-resampler==0.8.3.1-cyscaret==3.4.0) (2.0.0)
  Requirement already satisfied: retrying in /usr/local/lib/python3.10/dist-packages (from dash==2.9.0-splotly-resampler==0.8.3.1-cyscaret==3.4.0) (1.3.4)
  Requirement already satisfied: nest-asyncio in /usr/local/lib/python3.10/dist-packages (from dash==2.9.0-splotly-resampler==0.8.3.1-cyscaret==3.4.0) (1.6.0)
  Requirement already satisfied: jupyter-client in /usr/local/lib/python3.10/dist-packages (from ipykernel==6.5.1->ipywidgets==7.6.5->pycaret==3.4.0) (6.1.12)
  Requirement already satisfied: tornado>=4.2 in /usr/local/lib/python3.10/dist-packages (from ipykernel==6.5.1->ipywidgets==7.6.5->pycaret==3.4.0) (6.3.3)
  Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /usr/local/lib/python3.10/dist-packages (from jsonschema==2.6->nbformat==4.2.0->pycaret==3.4.0) (2023.12.1)
  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->pycaret==3.4.0) (0.8.4)
  Requirement already satisfied: platformdirs>=2.5 in /usr/local/lib/python3.10/dist-packages (from jupyter-core!=5.0.*,>=4.12->nbformat>=4.2.0->pycaret) (4.3.6)
  Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages (from patsy>=0.5.1->category-encoders==2.4.0->pycaret) (1.16.0)
  Requirement already satisfied: ptyprocess>=0.5 in /usr/local/lib/python3.10/dist-packages (from pexpect>4.3->ipython<=5.5.0->pycaret) (0.7.0)
  Requirement already satisfied: wcwidth in /usr/local/lib/python3.10/dist-packages (from prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0->ipython<=5.5.0->pycaret) (0.2.13)
  Requirement already satisfied: notebook>=4.4.1 in /usr/local/lib/python3.10/dist-packages (from widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (6.5.5)
  Requirement already satisfied: Flask<2.3.1,>=1.0.4 in /usr/local/lib/python3.10/dist-packages (from dash==2.9.0-splotly-resampler==0.8.3.1-cyscaret==3.4.0) (2.2.0)
  Requirement already satisfied: click>=8.0 in /usr/local/lib/python3.10/dist-packages (from Flask<2.3.1,>=1.0.4->dash==2.9.0-splotly-resampler==0.8.3.1-cyscaret==3.4.0) (8.1.7)
  Requirement already satisfied: psymc25,>=17 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (24.0.1)
  Requirement already satisfied: argon2-cffi in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (23.1.0)
  Requirement already satisfied: nbconvert<6 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (6.5.4)
  Requirement already satisfied: Send2Trash>=1.8.0 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (1.8.3)
  Requirement already satisfied: terminado>=0.8.3 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (0.18.1)
  Requirement already satisfied: prometheus-client in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (0.21.0)
  Requirement already satisfied: nbclassic>=0.4.7 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (1.1.0)
  Requirement already satisfied: notebook-shim>=0.2.3 in /usr/local/lib/python3.10/dist-packages (from nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (0.2.4)
  Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (4.9.4)
  Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (4.12.3)
  Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (6.1.0)
  Requirement already satisfied: defusedxml in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (0.7.1)
  Requirement already satisfied: entrypoints>0.2.2 in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (0.4)
  Requirement already satisfied: jupyterlab-pygments in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (0.3.0)
  Requirement already satisfied: mistune<2,>=0.8.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (0.8.4)
  Requirement already satisfied: nbclient>=0.5.0 in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (0.10.0)
  Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (1.5.1)
  Requirement already satisfied: tinycss2 in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (1.3.0)
  Requirement already satisfied: argon2-cffi-bindings in /usr/local/lib/python3.10/dist-packages (from argon2-cffi-notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (21.2.0)
  Requirement already satisfied: jupyter-server<3,>=1.8 in /usr/local/lib/python3.10/dist-packages (from notebook-shim>=0.2.3->nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (1.24.0)
  Requirement already satisfied: cffi>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from argon2-cffi-bindings->argon2-cffi-notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (1.17.1)
  Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-packages (from beautifulsoup4->nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (2.6)
  Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-packages (from bleach->nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (0.5.1)
  Requirement already satisfied: pycparser in /usr/local/lib/python3.10/dist-packages (from cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi-notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (2.22)
  Requirement already satisfied: anyio<4,>=3.1.0 in /usr/local/lib/python3.10/dist-packages (from jupyter-server<3,>=1.8->notebook-shim>=0.2.3->nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret==3.4.0) (3.7.1)
  Requirement already satisfied: websocket-client in /usr/local/lib/python3.10/dist-packages (from jupyter-server<3,>=1.8->notebook-shim>=0.2.3->nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret==3.4.0) (1.8.0)
```

```
!pip install -q --upgrade pip
!pip install -q pycaret
```

```
!pip install --upgrade pycaret
```

```
  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->pycaret) (3.0.11)
  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->pycaret) (3.3.2)
  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) (2024.8.30)
  Requirement already satisfied: attrs>=20.1.0 in /usr/local/lib/python3.10/dist-packages (from trio<0.25.0,>=0.22.0->pycaret==3.4.0) (24.2.0)
  Requirement already satisfied: sortedcontainers in /usr/local/lib/python3.10/dist-packages (from trio<0.25.0,>=0.22.0->pycaret) (2.4.0)
  Requirement already satisfied: outcome in /usr/local/lib/python3.10/dist-packages (from trio<0.25.0,>=0.22.0->pycaret) (1.3.0.post0)
  Requirement already satisfied: sniffio>=1.3.0 in /usr/local/lib/python3.10/dist-packages (from trio<0.25.0,>=0.22.0->pycaret==3.4.0) (1.3.1)
  Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-packages (from trio<0.25.0,>=0.22.0->pycaret==3.4.0) (1.2.2)
  Requirement already satisfied: scikit-base<0.10.0,>=0.6.1 in /usr/local/lib/python3.10/dist-packages (from sktime->pycaret==3.4.0) (0.9.0)
  Requirement already satisfied: dash==2.9.0-splotly-resampler==0.8.3.1-cyscaret==3.4.0) (2.2.5)
  Requirement already satisfied: werkzeug==2.3.0 in /usr/local/lib/python3.10/dist-packages (from dash==2.9.0-splotly-resampler==0.8.3.1-cyscaret==3.4.0) (2.0.0)
  Requirement already satisfied: dash-core-components==2.0.0 in /usr/local/lib/python3.10/dist-packages (from dash==2.9.0-splotly-resampler==0.8.3.1-cyscaret==3.4.0) (2.0.0)
  Requirement already satisfied: dash-table==5.0.0 in /usr/local/lib/python3.10/dist-packages (from dash==2.9.0-splotly-resampler==0.8.3.1-cyscaret==3.4.0) (5.0.0)
  Requirement already satisfied: dash-html-components==2.0.0 in /usr/local/lib/python3.10/dist-packages (from dash==2.9.0-splotly-resampler==0.8.3.1-cyscaret==3.4.0) (2.0.0)
  Requirement already satisfied: retrying in /usr/local/lib/python3.10/dist-packages (from dash==2.9.0-splotly-resampler==0.8.3.1-cyscaret==3.4.0) (1.3.4)
  Requirement already satisfied: nest-asyncio in /usr/local/lib/python3.10/dist-packages (from dash==2.9.0-splotly-resampler==0.8.3.1-cyscaret==3.4.0) (1.6.0)
  Requirement already satisfied: jupyter-client in /usr/local/lib/python3.10/dist-packages (from ipykernel==6.5.1->ipywidgets==7.6.5->pycaret) (6.1.12)
  Requirement already satisfied: tornado>=4.2 in /usr/local/lib/python3.10/dist-packages (from ipykernel==6.5.1->ipywidgets==7.6.5->pycaret==3.4.0) (6.3.3)
  Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /usr/local/lib/python3.10/dist-packages (from jsonschema==2.6->nbformat==4.2.0->pycaret==3.4.0) (2023.12.1)
  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->pycaret==3.4.0) (0.8.4)
  Requirement already satisfied: platformdirs>=2.5 in /usr/local/lib/python3.10/dist-packages (from jupyter-core!=5.0.*,>=4.12->nbformat>=4.2.0->pycaret) (4.3.6)
  Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages (from patsy>=0.5.1->category-encoders==2.4.0->pycaret) (1.16.0)
  Requirement already satisfied: ptyprocess>=0.5 in /usr/local/lib/python3.10/dist-packages (from pexpect>4.3->ipython<=5.5.0->pycaret) (0.7.0)
  Requirement already satisfied: wcwidth in /usr/local/lib/python3.10/dist-packages (from prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0->ipython<=5.5.0->pycaret==3.4.0) (0.2.13)
  Requirement already satisfied: notebook>=4.4.1 in /usr/local/lib/python3.10/dist-packages (from widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (6.5.5)
  Requirement already satisfied: Flask<2.3.1,>=1.0.4 in /usr/local/lib/python3.10/dist-packages (from dash==2.9.0-splotly-resampler==0.8.3.1-cyscaret==3.4.0) (2.2.5)
  Requirement already satisfied: click>=8.0 in /usr/local/lib/python3.10/dist-packages (from Flask<2.3.1,>=1.0.4->dash==2.9.0-splotly-resampler==0.8.3.1-cyscaret==3.4.0) (8.1.7)
  Requirement already satisfied: psymc25,>=17 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (24.0.1)
  Requirement already satisfied: argon2-cffi in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (23.1.0)
  Requirement already satisfied: nbconvert<6 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (6.5.4)
  Requirement already satisfied: Send2Trash>=1.8.0 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (1.8.3)
  Requirement already satisfied: terminado>=0.8.3 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (0.18.1)
  Requirement already satisfied: prometheus-client in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (0.21.0)
  Requirement already satisfied: nbclassic>=0.4.7 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (1.1.0)
  Requirement already satisfied: notebook-shim>=0.2.3 in /usr/local/lib/python3.10/dist-packages (from nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret==3.4.0) (0.2.4)
  Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (4.9.4)
  Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (4.12.3)
  Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (6.1.0)
  Requirement already satisfied: defusedxml in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (0.7.1)
  Requirement already satisfied: entrypoints>0.2.2 in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (0.4)
  Requirement already satisfied: jupyterlab-pygments in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (0.3.0)
  Requirement already satisfied: mistune<2,>=0.8.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (0.8.4)
  Requirement already satisfied: nbclient>=0.5.0 in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (0.10.0)
  Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (1.5.1)
  Requirement already satisfied: tinycss2 in /usr/local/lib/python3.10/dist-packages (from nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (1.3.0)
  Requirement already satisfied: argon2-cffi-bindings in /usr/local/lib/python3.10/dist-packages (from argon2-cffi-notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (21.2.0)
  Requirement already satisfied: jupyter-server<3,>=1.8 in /usr/local/lib/python3.10/dist-packages (from notebook-shim>=0.2.3->nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (1.24.0)
  Requirement already satisfied: cffi>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from argon2-cffi-bindings->argon2-cffi-notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (1.17.1)
  Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-packages (from beautifulsoup4->nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (2.6)
  Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-packages (from bleach->nbconvert<6->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret) (0.5.1)
  Requirement already satisfied: pycparser in /usr/local/lib/python3.10/dist-packages (from cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi-notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret==3.4.0) (2.22)
  Requirement already satisfied: anyio<4,>=3.1.0 in /usr/local/lib/python3.10/dist-packages (from jupyter-server<3,>=1.8->notebook-shim>=0.2.3->nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret==3.4.0) (3.7.1)
  Requirement already satisfied: websocket-client in /usr/local/lib/python3.10/dist-packages (from jupyter-server<3,>=1.8->notebook-shim>=0.2.3->nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension==3.6.0->ipywidgets==7.6.5->pycaret==3.4.0) (1.8.0)
```

```
# Import pycaret classification and init setup
from pycaret.classification import *
setup(data, target = 'Outcome', session_id = 123)
```

	Description	Value
0	Session id	123
1	Target	Outcome
2	Target type	Binary
3	Original data shape	(99, 22)
4	Transformed data shape	(99, 31)
5	Transformed train set shape	(69, 31)
6	Transformed test set shape	(30, 31)
7	Numeric features	12
8	Categorical features	9
9	Preprocess	True
10	Imputation type	simple
11	Numeric imputation	mean
12	Categorical imputation	mode
13	Maximum one-hot encoding	25
14	Encoding method	None
15	Fold Generator	StratifiedKFold
16	Fold Number	10
17	CPU Jobs	-1
18	Use GPU	False
19	Log Experiment	False
20	Experiment Name	clf-default-name
21	USI	4855

```
<pycaret.classification.oop.ClassificationExperiment at 0x7a0b82ba4a40>
```

```
!pip install --upgrade pandas
```

```
  Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (2.2.3)
  Requirement already satisfied: numpy>=1.22.4 in /usr/local/lib/python3.10/dist-packages (from pandas) (1.26.4)
  Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.10/dist-packages (from pandas) (2.8.2)
  Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas) (2024.2)
  Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.10/dist-packages (from pandas) (2024.1)
  Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
```

Compare Models

```
# compare baseline models
best = compare_models()
```

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
lightgbm	Light Gradient Boosting Machine	0.9286	0.3833	0.1000	0.1000	0.1000	nan	0.1000	0.4220
knn	K Neighbors Classifier	0.9143	0.2833	0.0000	0.0000	0.0000	nan	0.0000	0.1550
qda	Quadratic Discriminant Analysis	0.9143	0.3000	0.0000	0.0000	0.0000	nan	0.0000	0.1650
dummy	Dummy Classifier	0.9143	0.3000	0.0000	0.0000	0.0000	nan	0.0000	0.2700
svm	SVM - Linear Kernel	0.9000	0.3667	0.0000	0.0000	0.0000	nan	0.0000	0.1490
ridge	Ridge Classifier	0.8976	0.4500	0.0000	0.0000	0.0000	nan	0.0000	0.1500
rf	Random Forest Classifier	0.8976	0.3750	0.0000	0.0000	0.0000	nan	0.0000	0.3090
ada	Ada Boost Classifier	0.8976	0.5000	0.1000	0.1000	0.1000	nan	0.1000	0.3200
et	Extra Trees Classifier	0.8976	0.4250	0.0000	0.0000	0.0000	nan	0.0000	0.2710
xgboost	Extreme Gradient Boosting	0.8667	0.4000	0.0000	0.0000	0.0000	nan	-0.0167	0.1980
lr	Logistic Regression	0.8524	0.3333	0.2000	0.1500	0.1667	nan	0.1479	0.8130
gbc	Gradient Boosting Classifier	0.8524	0.3833	0.0000	0.0000	0.0000	nan	-0.0333	0.3990
lda	Linear Discriminant Analysis	0.8262	0.4500	0.1000	0.0333	0.0500	nan	0.0138	0.1460
dt	Decision Tree Classifier	0.7952	0.3417	0.2000	0.0833	0.1167	nan	0.0450	0.2720
nb	Naive Bayes	0.5643	0.4333	0.4000	0.1083	0.1700	0.0438	0.0589	0.2430

Create Model

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

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
Fold							
0	1.0000	0.0000	0.0000	0.0000	0.0000	nan	0.0000
1	1.0000	0.0000	0.0000	0.0000	0.0000	nan	0.0000
2	1.0000	0.0000	0.0000	0.0000	0.0000	nan	0.0000
3	0.8571	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.8571	0.3333	0.0000	0.0000	0.0000	0.0000	0.0000
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0.8571	0.1667	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.8571	0.3333	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.8571	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
9	1.0000	0.0000	0.0000	0.0000	0.0000	nan	0.0000
Mean	0.9286	0.3833	0.1000	0.1000	0.1000	nan	0.1000
Std	0.0714	0.4220	0.3000	0.3000	0.3000	nan	0.3000

```
# print model parameters
print(lightgbm)
```

```
LGBMClassifier(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, subsample=1.0,
               subsample_for_bin=200000, subsample_freq=0)
```

Tune Model

```
tuned_lightgbm = tune_model(lightgbm)
```

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
Fold							
0	1.0000	0.0000	0.0000	0.0000	0.0000	nan	0.0000
1	1.0000	0.0000	0.0000	0.0000	0.0000	nan	0.0000
2	1.0000	0.0000	0.0000	0.0000	0.0000	nan	0.0000
3	0.8571	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.8571	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.8571	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.8571	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.8571	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.8571	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
9	1.0000	0.0000	0.0000	0.0000	0.0000	nan	0.0000
Mean	0.9143	0.3000	0.0000	0.0000	0.0000	nan	0.0000
Std	0.0700	0.2449	0.0000	0.0000	0.0000	nan	0.0000

```
Fitting 10 folds for each of 10 candidates, totalling 100 fits
Original model was better than the tuned model, hence it will be returned. NOTE: The display metrics are for the tuned model (not the original one).
```

```
tuned_lightgbm, tuner = tune_model(lightgbm, return_tuner=True)
```

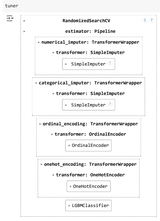
	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
Fold							
0	1.0000	0.0000	0.0000	0.0000	0.0000	nan	0.0000
1	1.0000	0.0000	0.0000	0.0000	0.0000	nan	0.0000
2	1.0000	0.0000	0.0000	0.0000	0.0000	nan	0.0000
3	0.8571	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.8571	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.8571	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.8571	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.8571	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.8571	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
9	1.0000	0.0000	0.0000	0.0000	0.0000	nan	0.0000
Mean	0.9143	0.3000	0.0000	0.0000	0.0000	nan	0.0000
Std	0.0700	0.2449	0.0000	0.0000	0.0000	nan	0.0000

```
Fitting 10 folds for each of 10 candidates, totalling 100 fits
Original model was better than the tuned model, hence it will be returned. NOTE: The display metrics are for the tuned model (not the original one).
```

```
tuned_lightgbm
```

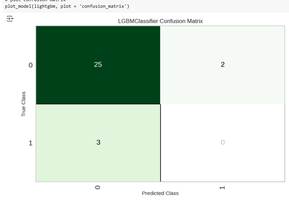
```
                        LGBMClassifier
LGBMClassifier(boosting_type='gbdt', class_weight=None, colsample_bytree=1.0,
               importance_type='split', learning_rate=0.1, max_depth=-1,
               min_child_samples=20, min_child_weight=0.001, min_split_gain=0.0,
               n_estimators=100, n_jobs=-1, num_leaves=31, objective=None,
               random_state=123, reg_alpha=0.0, reg_lambda=0.0, subsample=1.0,
               subsample_for_bin=200000, subsample_freq=0)
```

```
tuner
```

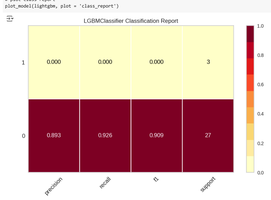


Analyze Model

```
# plot confusion matrix
plot_model(lightgbm, plot = 'confusion_matrix')
```

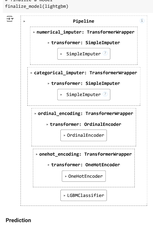



```
# plot class report
plot_model(lightgbm, plot = 'class_report')
```



Evaluate Model

```
# finalize a model
finalize_model(lightgbm)
```



Prediction

```
# predict on test set
holdout_pred = predict_model(lightgbm)
```

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	Light Gradient Boosting Machine	0.8333	0.4444	0.0000	0.0000	0.0000	-0.0870	-0.0891

```
# show predictions
holdout_pred.head()
```

```
    AGE  PackHistory  COPDSEVERITY  MWT1  MWT2  MWTBest  FEV1  FEV1PRED  FVC  FVCPRED  ...  copd  gender  smoking  Diabetes  muscular  hypertension  AtrialFib  Outcome  prediction_label  prediction_score
37   74          64.0      MODERATE 384.0 372.0    384.0  0.37      74.0 4.70    111  ...     2       1        2         0         1             0          0        0                 0            0.7810
44   73          59.0      MODERATE 405.0 405.0    405.0  2.43      75.0 5.37    125  ...     2       1        2         0         0             0          0        0                 0            0.7161
39   71          20.0          MILD 390.0 390.0    390.0  1.32      85.0 2.27    119  ...     1       0        2         1         0             0          0        0                 0            0.9831
72   75           5.0      MODERATE 462.0 423.0    462.0  2.43      79.0 4.33    107  ...     2       1        2         0         0             0          0        0                 0            0.9494
50   65          34.0      MODERATE 420.0 420.0    420.0  1.45      61.0 2.85    101  ...     2       0        1         0         1             0          0        0                 0            0.9712

5 rows × 24 columns
```

```
# copy data and drop Class variable
new_data = data.copy()
new_data.drop('Outcome', axis=1, inplace=True)
new_data.head()
```

```
   AGE  PackHistory  COPDSEVERITY  MWT1  MWT2  MWTBest  FEV1  FEV1PRED  FVC  FVCPRED  ...  HAD  SGRQ  AGEquartiles  copd  gender  smoking  Diabetes  muscular  hypertension  AtrialFib
0   77         60.0        SEVERE 120.0 120.0    120.0  1.21      36.0 2.40       98  ...  8.0 69.55             4     3       1        2         1         0             0          1
1   79         50.0      MODERATE 165.0 176.0    176.0  1.09      56.0 1.64       65  ... 21.0 44.24             4     2       0        2         1         0             0          1
2   80         11.0      MODERATE 201.0 180.0    201.0  1.52      68.0 2.30       86  ... 18.0 44.09             4     2       0        2         1         0             0          1
3   56         60.0 VERY SEVERE 210.0 210.0    210.0  0.47      14.0 1.14       27  ... 26.0 62.04             1     4       1        2         0         0             1          1
4   65         68.0        SEVERE 204.0 210.0    210.0  1.07      42.0 2.91       98  ... 18.0 75.56             1     3       1        2         0         1             1          0

5 rows × 21 columns
```

```
# predict model on new_data
predictions = predict_model(lightgbm, data = new_data)
predictions.head()
```

```
   AGE  PackHistory  COPDSEVERITY  MWT1  MWT2  MWTBest  FEV1  FEV1PRED  FVC  FVCPRED  ...  AGEquartiles  copd  gender  smoking  Diabetes  muscular  hypertension  AtrialFib  prediction_label  prediction_score
0   77         60.0        SEVERE 120.0 120.0    120.0  1.21      36.0 2.40       98  ...             4     3       1        2         1         0             0          1                 0            0.9080
1   79         50.0      MODERATE 165.0 176.0    176.0  1.09      56.0 1.64       65  ...             4     2       0        2         1         0             0          1                 0            0.9166
2   80         11.0      MODERATE 201.0 180.0    201.0  1.52      68.0 2.30       86  ...             4     2       0        2         1         0             0          1                 0            0.9997
3   56         60.0 VERY SEVERE 210.0 210.0    210.0  0.47      14.0 1.14       27  ...             1     4       1        2         0         0             1          1                 0            0.7810
4   65         68.0        SEVERE 204.0 210.0    210.0  1.07      42.0 2.91       98  ...             1     3       1        2         0         1             1          0                 0            0.7161

5 rows × 23 columns
```

Save Model

```
# save pipeline
save_model(lightgbm, 'E:\COPD_Model')
```

```
Transformation Pipeline and Model Successfully Saved
(Pipeline(memory=Memory(location=None),
          steps=[('numerical_imputer',
                  TransformerWrapper(exclude=None,
                                     include=['AGE', 'PackHistory', 'MWT1',
                                              'MWT2', 'MWTBest', 'FEV1',
                                              'FEV1PRED', 'FVC', 'FVCPRED',
                                              'CAT', 'HAD', 'SGRQ'],
                                     transformer=SimpleImputer(add_indicator=False,
                                                               copy=True,
                                                               fill_value=None,
                                                               keep_empty_features=False,
                                                               missing_values=nan,
                                                               strategy='mean'))),
                 ('categ...
                  LGBMClassifier(boosting_type='gbdt', class_weight=None,
                                 colsample_bytree=1.0, device_type='split',
                                 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, subsample=1.0,
                                 subsample_for_bin=200000, subsample_freq=0))],
          verbose=False),
 'E:\\COPD_Model.pkl')
```

```
# load pipeline
loaded_best_pipeline = load_model('E:\COPD_Model')
loaded_best_pipeline
```

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


Discussion

The Light Gradient Boosting Machine (LightGBM) model has predicted with precision at 83% accuracy.