Analytics for Hospitals Health-Care Data
TEAM ID: PNT2022TMID16326

IMPORT LIBRARIES

```python
import os
for dirname, _, filenames in os.walk('input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.colors import ListedColormap
import seaborn as sns
from warnings import filterwarnings
filterwarnings('ignore')
pd.options.display.max_columns = None
pd.options.display.max_rows = None
pd.options.display.float_format = '{:.6f}'.format
from sklearn.model_selection import train_test_split
import statsmodels
import statsmodels.api as sm
from sklearn.preprocessing import StandardScaler
from sklearn import metrics
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report
from sklearn.metrics import cohen_kappa_score
from sklearn.metrics import confusion_matrix
from sklearn.metrics import roc_curve
from sklearn.metrics import accuracy_score
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn import tree
from sklearn.model_selection import GridSearchCV
from sklearn.ensemble import AdaBoostClassifier,GradientBoostingClassifier
from catboost import CatBoostClassifier
from sklearn.feature_selection import RFE
plt.rcParams['figure.figsize'] = [15,8]
```

```python
train  = pd.read_csv("D:/HealthCare/train_data.csv")
test = pd.read_csv("D:/HealthCare/test_data.csv")
```

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

Out[4]:

	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Available Extra Rooms in Hospital	Department	Ward_Type	Ward_Facility_Code	Bed Grade	pati
0	1	8	c	3	Z	3	radiotherapy	R	F	2.000000	
1	2	2	c	5	Z	2	radiotherapy	S	F	2.000000	
2	3	10	e	1	X	2	anesthesia	S	E	2.000000	
3	4	26	b	2	Y	2	radiotherapy	R	D	2.000000	
4	5	26	b	2	Y	2	radiotherapy	S	D	2.000000	

```python
print(train.shape)
print(test.shape)
```

```
(318438, 18)
(137057, 17)
```

```python
train.dtypes
```

Out[6]:
```
case_id                 int64
Hospital_code           int64
Hospital_type_code      object
City Code Hospital      int64
```

```
Available Extra Rooms in Hospital     int64
Department                           object
Ward_Type                            object
Ward_Facility_Code                   object
Bed Grade                           float64
patientid                             int64
City_Code_Patient                   float64
Type of Admission                    object
Severity of Illness                  object
Visitors with Patient                int64
Age                                  object
Admission_Deposit                   float64
Stay                                 object
dtype: object
```

In [7]: 1 train.nunique()

Out[7]:
```
case_id                             318438
Hospital_code                           32
Hospital_type_code                       7
City_Code_Hospital                      11
Hospital_region_code                     3
Available Extra Rooms in Hospital       18
Department                               5
Ward_Type                                6
Ward_Facility_Code                       6
Bed Grade                                4
patientid                            92017
City_Code_Patient                       37
Type of Admission                        3
Severity of Illness                      3
Visitors with Patient                   28
Age                                     10
Admission_Deposit                     7300
Stay                                    11
dtype: int64
```

CHANGE DATA TYPES

In [8]: 1 train.duplicated().sum()

Out[8]: 0

In [9]:
```python
1  train['Hospital_code'] = train['Hospital_code'].astype(object)
2  train['City_Code_Hospital'] = train['City_Code_Hospital'].astype(object)
3  train['Available Extra Rooms in Hospital'] = train['Available Extra Rooms in Hospital'].astype(object)
4  train['Bed Grade'] = train['Bed Grade'].astype(object)
5  train['City_Code_Patient'] = train['City_Code_Patient'].astype(object)
```

In [10]: 1 train.dtypes

Out[10]:
```
case_id                              int64
Hospital_code                       object
Hospital_type_code                  object
City_Code_Hospital                  object
Hospital_region_code                object
Available Extra Rooms in Hospital   object
Department                          object
Ward_Type                           object
Ward_Facility_Code                  object
Bed Grade                           object
patientid                            int64
City_Code_Patient                   object
Type of Admission                   object
Severity of Illness                 object
Visitors with Patient                int64
Age                                 object
Admission_Deposit                  float64
Stay                                object
dtype: object
```

In [11]:
```python
1  test['Hospital_code'] = test['Hospital_code'].astype(object)
2  test['City_Code_Hospital'] = test['City_Code_Hospital'].astype(object)
3  test['Available Extra Rooms in Hospital'] = test['Available Extra Rooms in Hospital'].astype(object)
```

REMOVE INSIGNIFICANT VARIABLES

```
In [11]:    1  test['Hospital_code'] = test['Hospital_code'].astype(object)
            2  test['City_Code_Hospital'] = test['City_Code_Hospital'].astype(object)
            3  test['Available Extra Rooms in Hospital'] = test['Available Extra Rooms in Hospital'].astype(object)
            4  test['Bed Grade'] = test['Bed Grade'].astype(object)
            5  test['City_Code_Patient'] = test['City_Code_Patient'].astype(object)
```

```
In [12]:    1  test.dtypes
```

```
Out[12]: case_id                               int64
         Hospital_code                         object
         Hospital_type_code                    object
         City_Code_Hospital                    object
         Hospital_region_code                  object
         Available Extra Rooms in Hospital     object
         Department                            object
         Ward_Type                             object
         Ward_Facility_Code                    object
         Bed Grade                             object
         patientid                             int64
         City_Code_Patient                     object
         Type of Admission                     object
         Severity of Illness                   object
         Visitors with Patient                 int64
         Age                                   object
         Admission_Deposit                     float64
         dtype: object
```

```
In [13]:    1  train.drop(['case_id', 'patientid'], axis=1, inplace=True)
```

```
In [14]:    1  test.drop(['case_id', 'patientid'], axis=1, inplace=True)
```

```
In [15]:    1  train['Stay'] .replace ('More than 100 Days', '100+', inplace=True)
```

```
In [16]:    1  plt.figure(figsize=(15,8))
```

```
In [16]:    1  plt.figure(figsize=(15,8))
            2
            3  train.drop('Stay', axis = 1).hist()
            4  plt.tight_layout()
            5  plt.show()
            6
            7  print('Skewness:')
            8  train.drop('Stay', axis = 1).skew()
```

```
<Figure size 1500x800 with 0 Axes>
```

Skewness:

Out[16]: Hospital_code   -0.280783
         City_Code_Hospital   0.538809
         Available Extra Rooms in Hospital   0.971930
         Bed Grade   0.051754