In [11]:
```
import numpy as np
import pandas as pd

from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.model_selection import train_test_split

from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from sklearn.neural_network import MLPClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import AdaBoostClassifier, BaggingClassifier, Gradier
```

In [ ]:

In [ ]:

In [12]:
```
data = pd.read_csv('database_IND.csv')
```

In [ ]:

In [ ]:

In [ ]:

In [13]:
```
data
```

Out[13]:

	country	country_long	name	gppd_idnr	capacity_mw	latitude	longitude	prin
0	IND	India	ACME Solar Tower	WRI1020239	2.5	28.1839	73.2407	
1	IND	India	ADITYA CEMENT WORKS	WRI1019881	98.0	24.7663	74.6090	
2	IND	India	AES Saurashtra Windfarms	WRI1026669	39.2	21.9038	69.3732	
3	IND	India	AGARTALA GT	IND0000001	135.0	23.8712	91.3602	
4	IND	India	AKALTARA TPP	IND0000002	1800.0	21.9603	82.4091	
...	...	...	...	...	...	...	...	...
902	IND	India	YERMARUS TPP	IND0000513	1600.0	16.2949	77.3568	
903	IND	India	Yelesandra Solar Power Plant	WRI1026222	3.0	12.8932	78.1654	
904	IND	India	Yelisirur wind power project	WRI1026776	25.5	15.2758	75.5811	
905	IND	India	ZAWAR MINES	WRI1019901	80.0	24.3500	73.7477	
906	IND	India	iEnergy Theni Wind Farm	WRI1026761	16.5	9.9344	77.4768	

907 rows × 27 columns

In [ ]:
```

```

In [ ]:
```

```

In [14]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 907 entries, 0 to 906
Data columns (total 27 columns):
 #   Column                   Non-Null Count  Dtype  
---  ------                   --------------  -----  
 0   country                  907 non-null    object 
 1   country_long             907 non-null    object 
 2   name                     907 non-null    object 
 3   gppd_idnr                907 non-null    object 
 4   capacity_mw              907 non-null    float64
 5   latitude                 861 non-null    float64
 6   longitude                861 non-null    float64
 7   primary_fuel             907 non-null    object 
 8   other_fuel1              198 non-null    object 
 9   other_fuel2              1 non-null      object 
 10  other_fuel3              0 non-null      float64
 11  commissioning_year       527 non-null    float64
 12  owner                    342 non-null    object 
 13  source                   907 non-null    object 
 14  url                      907 non-null    object 
 15  geolocation_source       888 non-null    object 
 16  wepp_id                  0 non-null      float64
 17  year_of_capacity_data    519 non-null    float64
 18  generation_gwh_2013      0 non-null      float64
 19  generation_gwh_2014      398 non-null    float64
 20  generation_gwh_2015      422 non-null    float64
 21  generation_gwh_2016      434 non-null    float64
 22  generation_gwh_2017      440 non-null    float64
 23  generation_gwh_2018      448 non-null    float64
 24  generation_gwh_2019      0 non-null      float64
 25  generation_data_source   449 non-null    object 
 26  estimated_generation_gwh  0 non-null      float64
dtypes: float64(15), object(12)
memory usage: 191.4+ KB
```

In [ ]:
```

```

In [ ]:
```

```

In [15]:
```
data.isna().mean()
```

Out[15]:
```
country                    0.000000
country_long               0.000000
name                       0.000000
gppd_idnr                  0.000000
capacity_mw                0.000000
latitude                   0.050717
longitude                  0.050717
primary_fuel               0.000000
other_fuel1                0.781698
other_fuel2                0.998897
other_fuel3                1.000000
commissioning_year         0.418964
owner                      0.622933
source                     0.000000
url                        0.000000
geolocation_source         0.020948
wepp_id                    1.000000
year_of_capacity_data      0.427784
generation_gwh_2013        1.000000
generation_gwh_2014        0.561191
generation_gwh_2015        0.534730
generation_gwh_2016        0.521499
generation_gwh_2017        0.514884
generation_gwh_2018        0.506064
generation_gwh_2019        1.000000
generation_data_source     0.504961
estimated_generation_gwh   1.000000
dtype: float64
```

In [ ]:
```

```

In [16]:
```
data.shape
```

Out[16]:
```
(907, 27)
```

In [ ]:
```

```

In [17]:
```
print("Total missing values:", data.isna().sum().sum())
```

```
Total missing values: 10445
```

In [ ]:
```

```

In [18]:
```
data
```

Out[18]:

	country	country_long	name	gppd_idnr	capacity_mw	latitude	longitude	prin
0	IND	India	ACME Solar Tower	WRI1020239	2.5	28.1839	73.2407	
1	IND	India	ADITYA CEMENT WORKS	WRI1019881	98.0	24.7663	74.6090	
2	IND	India	AES Saurashtra Windfarms	WRI1026669	39.2	21.9038	69.3732	
3	IND	India	AGARTALA GT	IND0000001	135.0	23.8712	91.3602	
4	IND	India	AKALTARA TPP	IND0000002	1800.0	21.9603	82.4091	
...	...	...	...	...	...	...	...	...
902	IND	India	YERMARUS TPP	IND0000513	1600.0	16.2949	77.3568	
903	IND	India	Yelesandra Solar Power Plant	WRI1026222	3.0	12.8932	78.1654	
904	IND	India	Yelisirur wind power project	WRI1026776	25.5	15.2758	75.5811	
905	IND	India	ZAWAR MINES	WRI1019901	80.0	24.3500	73.7477	
906	IND	India	iEnergy Theni Wind Farm	WRI1026761	16.5	9.9344	77.4768	

907 rows × 27 columns

In [ ]:
```

```

In [10]:
```
data.head
```

Out[10]:
```
<bound method NDFrame.head of     country country_long
name   gppd_idnr  \
0       IND        India              ACME Solar Tower  WRI1020239
1       IND        India           ADITYA CEMENT WORKS  WRI1019881
2       IND        India      AES Saurashtra Windfarms  WRI1026669
3       IND        India                   AGARTALA GT  IND0000001
4       IND        India                 AKALTARA TPP   IND0000002
..      ...          ...                           ...         ...
902     IND        India                 YERMARUS TPP   IND0000513
903     IND        India  Yelesandra Solar Power Plant  WRI1026222
904     IND        India  Yelisirur wind power project  WRI1026776
905     IND        India                   ZAWAR MINES  WRI1019901
906     IND        India       iEnergy Theni Wind Farm  WRI1026761

     capacity_mw  latitude  longitude primary_fuel other_fuel1 other_fuel
2  \
0            2.5   28.1839    73.2407        Solar         NaN         Na
N
1           98.0   24.7663    74.6090         Coal         NaN         Na
N
2           39.2   21.9038    69.3732         Wind         NaN         Na
N
3          135.0   23.8712    91.3602          Gas         NaN         Na
N
4         1800.0   21.9603    82.4091         Coal         Oil         Na
N
..           ...       ...        ...          ...         ...
...
902       1600.0   16.2949    77.3568         Coal         Oil         Na
N
903          3.0   12.8932    78.1654        Solar         NaN         Na
N
904         25.5   15.2758    75.5811         Wind         NaN         Na
N
905         80.0   24.3500    73.7477         Coal         NaN         Na
N
906         16.5    9.9344    77.4768         Wind         NaN         Na
N

     ...  year_of_capacity_data  generation_gwh_2013 generation_gwh_2014
\
0    ...                    NaN                  NaN                 NaN
1    ...                    NaN                  NaN                 NaN
2    ...                    NaN                  NaN                 NaN
3    ...                 2019.0                  NaN          617.789264
4    ...                 2019.0                  NaN         3035.550000
..   ...                    ...                  ...                 ...
902  ...                 2019.0                  NaN                 NaN
903  ...                    NaN                  NaN                 NaN
904  ...                    NaN                  NaN                 NaN
905  ...                    NaN                  NaN                 NaN
906  ...                    NaN                  NaN                 NaN

    generation_gwh_2015 generation_gwh_2016 generation_gwh_2017  \
0                   NaN                 NaN                 NaN
1                   NaN                 NaN                 NaN
2                   NaN                 NaN                 NaN
```

```
3            843.747000          886.004428          663.774500
4           5916.370000         6243.000000         5385.579736
..                  ...                 ...                 ...
902            0.994875          233.596650          865.400000
903                 NaN                 NaN                 NaN
904                 NaN                 NaN                 NaN
905                 NaN                 NaN                 NaN
906                 NaN                 NaN                 NaN

     generation_gwh_2018  generation_gwh_2019         generation_data_sou
rce  \
0                    NaN                  NaN
NaN
1                    NaN                  NaN
NaN
2                    NaN                  NaN
NaN
3             626.239128                  NaN  Central Electricity Author
ity
4            7279.000000                  NaN  Central Electricity Author
ity
..                   ...                  ...
...
902           686.500000                  NaN  Central Electricity Author
ity
903                  NaN                  NaN
NaN
904                  NaN                  NaN
NaN
905                  NaN                  NaN
NaN
906                  NaN                  NaN
NaN

     estimated_generation_gwh
0                         NaN
1                         NaN
2                         NaN
3                         NaN
4                         NaN
..                        ...
902                       NaN
903                       NaN
904                       NaN
905                       NaN
906                       NaN

[907 rows x 27 columns]>
```

In [ ]:

In [ ]:

In [ ]:

In [ ]: