

```
In [1]: import warnings
warnings.filterwarnings('ignore')
```

```
In [2]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [3]: from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
```

```
In [4]: from imblearn import under_sampling, over_sampling
from sklearn.linear_model import LogisticRegression
```

```
In [5]: from sklearn.ensemble import AdaBoostClassifier
from sklearn.metrics import classification_report, confusion_matrix
```

```
In [6]: from sklearn.metrics import accuracy_score
from sklearn.metrics import roc_auc_score
```

```
In [11]: data = pd.read_csv("Desktop\Solar Business\solar features.csv")
```

```
In [16]: data.head()
print('There are', str(len(data)), 'rows of data in this dataset')
print('There are', str(data.shape[1]), 'features in this dataset')
data.head(15)
```

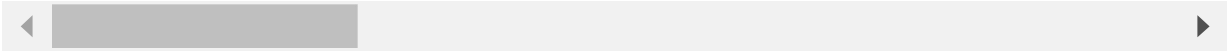
There are 169 rows of data in this dataset
There are 29 features in this dataset

```
Out[16]:
```

	prefecture_code	post_code	city	plant_name	founding_year	initial_funding_year	initial
0	SH	4380112	Iwata	Shimonobe	1/1/2023	1/2/2023	
1	SH	4380112	Iwata	Shimonobe	1/2/2023	1/3/2023	
2	SH	4380112	Iwata	Shimonobe	1/3/2023	1/4/2023	
3	SH	4380112	Iwata	Shimonobe	1/4/2023	1/5/2023	
4	SH	4380112	Iwata	Shimonobe	1/5/2023	1/6/2023	
5	SH	4350022	Hamamatsu	Tsurumi	1/6/2023	1/7/2023	
6	SH	4350022	Hamamatsu	Tsurumi	1/7/2023	1/8/2023	
7	SH	4350022	Hamamatsu	Tsurumi	1/8/2023	1/9/2023	
8	SH	4350022	Hamamatsu	Tsurumi	1/9/2023	1/10/2023	
9	SH	4350022	Hamamatsu	Tsurumi	1/10/2023	1/11/2023	
10	SH	4358560	Hamamatsu	Furukawa	1/11/2023	1/12/2023	
11	SH	4358560	Hamamatsu	Furukawa	1/12/2023	1/1/2024	

	prefecture_code	post_code	city	plant_name	founding_year	initial_funding_year	initial_funding_total_yen
12	SH	4358560	Hamamatsu	Furukawa	1/1/2024	1/2/2024	
13	SH	4358560	Hamamatsu	Furukawa	1/2/2024	1/3/2024	
14	SH	4358560	Hamamatsu	Furukawa	1/3/2024	1/4/2024	

15 rows × 29 columns



In [13]:

```
#Check Dataset
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 169 entries, 0 to 168
Data columns (total 29 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   prefecture_code                       169 non-null    object
1   post_code                             169 non-null    int64
2   city                                  169 non-null    object
3   plant_name                            169 non-null    object
4   founding_year                         169 non-null    object
5   initial_funding_year                  169 non-null    object
6   initial_funding_total_yen             169 non-null    object
7   intial_funding_age                    169 non-null    int64
8   milestones_achieved                   169 non-null    int64
9   first_funding_age                     169 non-null    int64
10  second_funding_age                    136 non-null    float64
11  has_collaboration                     169 non-null    int64
12  category_code                         169 non-null    object
13  is_automotive                         169 non-null    int64
14  is_energy                             169 non-null    int64
15  is_healthcare                         169 non-null    int64
16  has_vc                                169 non-null    int64
17  has_new_employees                     169 non-null    int64
18  employee_count                        169 non-null    int64
19  business_model_type                   169 non-null    int64
20  number_collaboration                  169 non-null    int64
21  competitor_count                      169 non-null    int64
22  research_team_size                    169 non-null    int64
23  time_to_market                        169 non-null    int64
24  industry_knowledge                    169 non-null    int64
25  social_value_creation                  169 non-null    int64
26  market_estimation                     169 non-null    int64
27  has_domain                            169 non-null    int64
28  target_marketsize                     169 non-null    int64
dtypes: float64(1), int64(21), object(7)
memory usage: 38.4+ KB
```

In [14]:

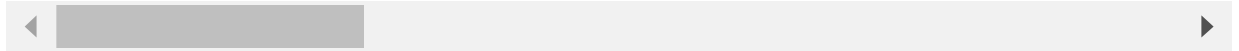
```
data.describe()
```

Out[14]:

	post_code	intial_funding_age	milestones_achieved	first_funding_age	second_funding_age
count	1.690000e+02	169.000000	169.000000	169.000000	136.000000
mean	4.363305e+06	20.698225	2.023669	86.295858	127.580882
std	1.281936e+04	1.357540	1.184989	19.932506	23.141611

	post_code	intial_funding_age	milestones_achieved	first_funding_age	second_funding_age
min	4.350022e+06	18.000000	0.000000	50.000000	90.000000
25%	4.350022e+06	20.000000	1.000000	70.000000	110.000000
50%	4.358560e+06	21.000000	2.000000	85.000000	123.500000
75%	4.380112e+06	21.000000	3.000000	100.000000	145.000000
max	4.380112e+06	25.000000	5.000000	150.000000	185.000000

8 rows × 22 columns



In [17]: `data['has_collaboration'].value_counts()`

Out[17]:
 0 90
 1 79
 Name: has_collaboration, dtype: int64

In [18]: `data['business_model_type'].value_counts()`

Out[18]:
 1 59
 2 57
 0 53
 Name: business_model_type, dtype: int64

In [19]: `data['is_automotive'].value_counts()`

Out[19]:
 0 117
 1 52
 Name: is_automotive, dtype: int64

In [20]: `data['is_energy'].value_counts()`

Out[20]:
 0 113
 1 56
 Name: is_energy, dtype: int64

In [21]: `data['is_healthcare'].value_counts()`

Out[21]:
 0 108
 1 61
 Name: is_healthcare, dtype: int64

In [23]: `data['industry_knowledge'].value_counts()`

Out[23]:
 0 87
 1 82
 Name: industry_knowledge, dtype: int64

In []: