almabetter-ml-assignment

August 12, 2024

[3]:
```
import pandas as pd
```

[5]:
```
df=pd.read_csv("estadistical.csv")
```

[7]:
```
df
```

[7]:

	Status of existing checking account	Duration in month	Credit history	\
0	A11	6	A34	
1	A12	48	A32	
2	A14	12	A34	
3	A11	42	A32	
4	A11	24	A33	
..	…	…	…	
995	A14	12	A32	
996	A11	30	A32	
997	A14	12	A32	
998	A11	45	A32	
999	A12	45	A34	

	Purpose	Credit amount	Savings account/bonds	Present employment since	\
0	A43	1169	A65	A75	
1	A43	5951	A61	A73	
2	A46	2096	A61	A74	
3	A42	7882	A61	A74	
4	A40	4870	A61	A73	
..	…	…	…	…	
995	A42	1736	A61	A74	
996	A41	3857	A61	A73	
997	A43	804	A61	A75	
998	A43	1845	A61	A73	
999	A41	4576	A62	A71	

	Installment rate in percentage of disposable income	\
0	4	
1	2	
2	2	
3	2	

```
4                                                     3
..                                                  …
995                                                   3
996                                                   4
997                                                   4
998                                                   4
999                                                   3

    Personal status and sex Other debtors / guarantors  …  Property  \
0                       A93                       A101  …      A121
1                       A92                       A101  …      A121
2                       A93                       A101  …      A121
3                       A93                       A103  …      A122
4                       A93                       A101  …      A124
..                      …                        …  …       …
995                     A92                       A101  …      A121
996                     A91                       A101  …      A122
997                     A93                       A101  …      A123
998                     A93                       A101  …      A124
999                     A93                       A101  …      A123

    Age in years  Other installment plans  Housing  \
0             67                     A143    A152
1             22                     A143    A152
2             49                     A143    A152
3             45                     A143    A153
4             53                     A143    A153
..           …                      …     …
995           31                     A143    A152
996           40                     A143    A152
997           38                     A143    A152
998           23                     A143    A153
999           27                     A143    A152

    Number of existing credits at this bank   Job  \
0                                         2  A173
1                                         1  A173
2                                         1  A172
3                                         1  A173
4                                         2  A173
..                                       …   …
995                                       1  A172
996                                       1  A174
997                                       1  A173
998                                       1  A173
999                                       1  A173
```

```
    Number of people being liable to provide maintenance for  Telephone  \
0                                                          1        A192
1                                                          1        A191
2                                                          2        A191
3                                                          2        A191
4                                                          2        A191
..                                                       ...         ...
995                                                        1        A191
996                                                        1        A192
997                                                        1        A191
998                                                        1        A192
999                                                        1        A191

    foreign worker Receive/ Not receive credit
0             A201                          1
1             A201                          2
2             A201                          1
3             A201                          1
4             A201                          2
..             ...                        ...
995           A201                          1
996           A201                          1
997           A201                          1
998           A201                          2
999           A201                          1

[1000 rows x 21 columns]
```

[9]:
```
df.shape
```

[9]:
```
(1000, 21)
```

[11]:
```
df.isna().sum()
```

[11]:
```
Status of existing checking account                        0
Duration in month                                          0
Credit history                                             0
Purpose                                                    0
Credit amount                                              0
Savings account/bonds                                      0
Present employment since                                   0
Installment rate in percentage of disposable income        0
Personal status and sex                                    0
Other debtors / guarantors                                 0
Present residence since                                    0
Property                                                   0
Age in years                                               0
```

```
Other installment plans                                    0
Housing                                                    0
Number of existing credits at this bank                    0
Job                                                        0
Number of people being liable to provide maintenance for   0
Telephone                                                  0
foreign worker                                             0
Receive/ Not receive credit                                0
dtype: int64
```

[13]:
```
df.describe()
```

[13]:
```
       Duration in month  Credit amount  \
count        1000.000000    1000.000000
mean           20.903000    3271.258000
std            12.058814    2822.736876
min             4.000000     250.000000
25%            12.000000    1365.500000
50%            18.000000    2319.500000
75%            24.000000    3972.250000
max            72.000000   18424.000000

       Installment rate in percentage of disposable income  \
count                                          1000.000000
mean                                              2.973000
std                                               1.118715
min                                               1.000000
25%                                               2.000000
50%                                               3.000000
75%                                               4.000000
max                                               4.000000

       Present residence since  Age in years  \
count              1000.000000   1000.000000
mean                  2.845000     35.546000
std                   1.103718     11.375469
min                   1.000000     19.000000
25%                   2.000000     27.000000
50%                   3.000000     33.000000
75%                   4.000000     42.000000
max                   4.000000     75.000000

       Number of existing credits at this bank  \
count                                1000.000000
mean                                    1.407000
std                                     0.577654
min                                     1.000000
```

```
25%                                    1.000000
50%                                    1.000000
75%                                    2.000000
max                                    4.000000

       Number of people being liable to provide maintenance for  \
count                                          1000.000000
mean                                              1.155000
std                                               0.362086
min                                               1.000000
25%                                               1.000000
50%                                               1.000000
75%                                               1.000000
max                                               2.000000

       Receive/ Not receive credit
count                   1000.000000
mean                       1.300000
std                        0.458487
min                        1.000000
25%                        1.000000
50%                        1.000000
75%                        2.000000
max                        2.000000
```

[15]:
```
print(df.describe(include=['object']))
```

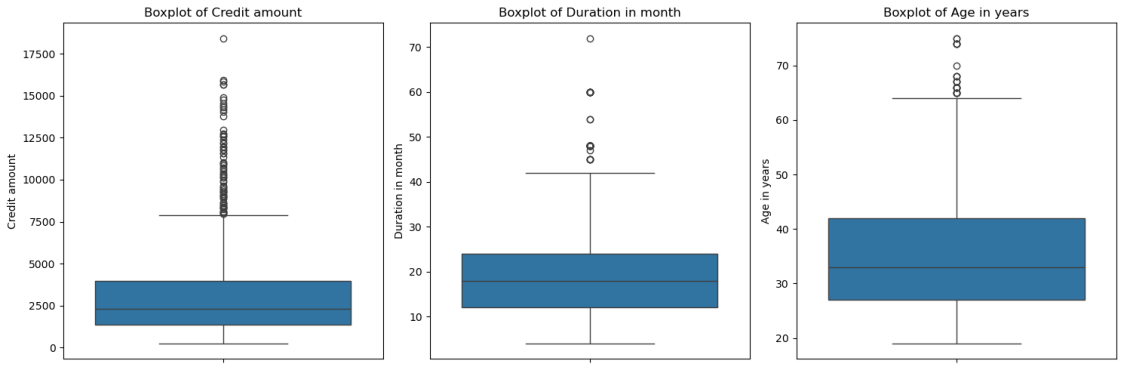
```
       Status of existing checking account Credit history Purpose  \
count                                 1000           1000    1000
unique                                   4              5      10
top                                    A14            A32    A43
freq                                   394            530    280

       Savings account/bonds Present employment since Personal status and sex  \
count                   1000                    1000                     1000
unique                     5                       5                        4
top                      A61                     A73                      A93
freq                     603                     339                      548

       Other debtors / guarantors Property Other installment plans  Housing  \
count                        1000     1000                    1000     1000
unique                          3        4                       3        3
top                          A101     A123                    A143     A152
freq                          907      332                     814      713

         Job Telephone foreign worker
count   1000      1000           1000
```

```
unique      4          2              2
top      A173       A191           A201
freq      630        596            963
```

[17]:
```
df.duplicated().sum()
```

[17]: 0

[19]:
```
#checking outliers
import matplotlib.pyplot as plt
import seaborn as sns

# Select numerical columns for outlier detection
numerical_columns = ['Credit amount', 'Duration in month', 'Age in years']

# Create boxplots for each numerical column
plt.figure(figsize=(15, 5))

for i, column in enumerate(numerical_columns):
    plt.subplot(1, len(numerical_columns), i + 1)
    sns.boxplot(df[column])
    plt.title(f'Boxplot of {column}')

plt.tight_layout()
plt.show()
```



[20]:
```
#IQR
# Function to detect outliers based on IQR
def detect_outliers_iqr(df, column):
    Q1 = df[column].quantile(0.25)
    Q3 = df[column].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
```

```
    return df[(df[column] < lower_bound) | (df[column] > upper_bound)]

# Apply the IQR method to each numerical column
for column in numerical_columns:
    outliers = detect_outliers_iqr(df, column)
    print(f"Outliers in {column}:")
    print(outliers)
```

```
Outliers in Credit amount:
    Status of existing checking account  Duration in month Credit history  \
5                                   A14                36            A32
17                                  A11                30            A30
18                                  A12                24            A32
57                                  A14                36            A34
63                                  A12                48            A30
..                                  …                  …             …
927                                 A11                48            A32
945                                 A12                48            A30
953                                 A14                36            A32
980                                 A12                30            A34
983                                 A11                36            A32

    Purpose  Credit amount Savings account/bonds Present employment since  \
5       A46           9055                  A65                       A73
17      A49           8072                  A65                       A72
18      A41          12579                  A61                       A75
57      A43           9566                  A61                       A73
63      A49          14421                  A61                       A73
..      …            …                      …                         …
927     A41          10297                  A61                       A74
945     A40           8358                  A63                       A72
953     A42          10974                  A61                       A71
980     A42           8386                  A61                       A74
983     A41           8229                  A61                       A73

     Installment rate in percentage of disposable income  \
5                                                    2
17                                                   2
18                                                   4
57                                                   2
63                                                   2
..                                                   …
927                                                  4
945                                                  1
953                                                  4
980                                                  2
983                                                  2
```

```
    Personal status and sex Other debtors / guarantors  …  Property  \
5                       A93                       A101  …      A124
17                      A93                       A101  …      A123
18                      A92                       A101  …      A124
57                      A92                       A101  …      A123
63                      A93                       A101  …      A123
..                      …                        …  …       …
927                     A93                       A101  …      A124
945                     A92                       A101  …      A123
953                     A92                       A101  …      A123
980                     A93                       A101  …      A122
983                     A93                       A101  …      A122

    Age in years  Other installment plans  Housing  \
5             35                    A143    A153
17            25                    A141    A152
18            44                    A143    A153
57            31                    A142    A152
63            25                    A143    A152
..           …                     …     …
927           39                    A142    A153
945           30                    A143    A152
953           26                    A143    A152
980           49                    A143    A152
983           26                    A143    A152

    Number of existing credits at this bank   Job  \
5                                         1  A172
17                                        3  A173
18                                        1  A174
57                                        2  A173
63                                        1  A173
..                                       …   …
927                                       3  A173
945                                       2  A173
953                                       2  A174
980                                       1  A173
983                                       1  A173

    Number of people being liable to provide maintenance for  Telephone  \
5                                                         2        A192
17                                                        1        A191
18                                                        1        A192
57                                                        1        A191
63                                                        1        A192
..                                                       …           …
927                                                       2        A192
```

```
945                                                  1             A191
953                                                  1             A192
980                                                  1             A191
983                                                  2             A191

    foreign worker Receive/ Not receive credit
5             A201                              1
17            A201                              1
18            A201                              2
57            A201                              1
63            A201                              2
..             …                              …
927           A201                              2
945           A201                              1
953           A201                              2
980           A201                              2
983           A201                              2

[72 rows x 21 columns]
Outliers in Duration in month:
    Status of existing checking account  Duration in month Credit history  \
1                                   A12                48            A32
11                                  A11                48            A32
29                                  A11                60            A33
35                                  A12                45            A34
36                                  A14                48            A34
..                                  …                  …             …
945                                 A12                48            A30
973                                 A11                60            A32
981                                 A14                48            A32
998                                 A11                45            A32
999                                 A12                45            A34

    Purpose  Credit amount Savings account/bonds Present employment since  \
1       A43           5951                  A61                       A73
11      A49           4308                  A61                       A72
29      A49           6836                  A61                       A75
35      A43           4746                  A61                       A72
36      A46           6110                  A61                       A73
..      …             …                     …                         …
945     A40           8358                  A63                       A72
973     A49           7297                  A61                       A75
981     A49           4844                  A61                       A71
998     A43           1845                  A61                       A73
999     A41           4576                  A62                       A71

     Installment rate in percentage of disposable income  \
1                                                    2
```

```
11                                                    3
29                                                    3
35                                                    4
36                                                    1
..                                                  …
945                                                   1
973                                                   4
981                                                   3
998                                                   4
999                                                   3

    Personal status and sex Other debtors / guarantors  …  Property  \
1                       A92                       A101  …      A121
11                      A92                       A101  …      A122
29                      A93                       A101  …      A124
35                      A93                       A101  …      A122
36                      A93                       A101  …      A124
..                      …                        …  …       …
945                     A92                       A101  …      A123
973                     A93                       A102  …      A124
981                     A93                       A101  …      A123
998                     A93                       A101  …      A124
999                     A93                       A101  …      A123

    Age in years  Other installment plans  Housing  \
1             22                    A143    A152
11            24                    A143    A151
29            63                    A143    A152
35            25                    A143    A152
36            31                    A141    A153
..           …                     …     …
945           30                    A143    A152
973           36                    A143    A151
981           33                    A141    A151
998           23                    A143    A153
999           27                    A143    A152

    Number of existing credits at this bank   Job  \
1                                         1  A173
11                                        1  A173
29                                        2  A173
35                                        2  A172
36                                        1  A173
..                                      …   …
945                                       2  A173
973                                       1  A173
981                                       1  A174
998                                       1  A173
```

```
999                                      1  A173

    Number of people being liable to provide maintenance for  Telephone  \
1                                                         1             A191
11                                                        1             A191
29                                                        1             A192
35                                                        1             A191
36                                                        1             A192
..                                                      …               …
945                                                       1             A191
973                                                       1             A191
981                                                       1             A192
998                                                       1             A192
999                                                       1             A191

    foreign worker Receive/ Not receive credit
1             A201                            2
11            A201                            2
29            A201                            2
35            A201                            2
36            A201                            1
..             …                            …
945           A201                            1
973           A201                            2
981           A201                            2
998           A201                            2
999           A201                            1

[70 rows x 21 columns]
Outliers in Age in years:
    Status of existing checking account  Duration in month Credit history  \
0                                   A11                  6            A34
75                                  A11                 12            A34
137                                 A12                 12            A32
163                                 A12                 10            A32
179                                 A11                 21            A34
186                                 A12                  9            A31
187                                 A12                 16            A34
213                                 A13                 30            A33
330                                 A11                 24            A34
430                                 A14                  5            A32
438                                 A11                 42            A34
536                                 A11                  6            A32
554                                 A12                  9            A32
606                                 A14                 24            A34
624                                 A11                 18            A32
723                                 A12                  9            A32
756                                 A13                  6            A34
```

774		A13	12	A34
779		A12	18	A32
807		A14	12	A34
846		A14	18	A32
883		A14	18	A34
917		A11	6	A32

	Purpose	Credit amount	Savings account/bonds	Present employment since	\
0	A43	1169	A65	A75	
75	A41	1526	A61	A75	
137	A43	766	A63	A73	
163	A40	7308	A61	A71	
179	A40	571	A61	A75	
186	A41	5129	A61	A75	
187	A40	1175	A61	A71	
213	A49	1908	A61	A75	
330	A41	6615	A61	A71	
430	A49	3448	A61	A74	
438	A45	3394	A61	A71	
536	A40	1374	A65	A71	
554	A46	1199	A61	A74	
606	A49	4526	A61	A73	
624	A43	2600	A61	A73	
723	A43	790	A63	A73	
756	A40	1299	A61	A73	
774	A40	1480	A63	A71	
779	A45	3872	A61	A71	
807	A43	930	A65	A75	
846	A40	6761	A65	A73	
883	A43	1098	A61	A71	
917	A40	14896	A61	A75	

	Installment rate in percentage of disposable income	\
0	4	
75	4	
137	4	
163	2	
179	4	
186	2	
187	2	
213	4	
330	2	
430	1	
438	4	
536	4	
554	4	
606	3	
624	4	

```
723                                                 4
756                                                 1
774                                                 2
779                                                 2
807                                                 4
846                                                 2
883                                                 4
917                                                 1

    Personal status and sex Other debtors / guarantors  …  Property  \
0                       A93                       A101  …      A121
75                      A93                       A101  …      A124
137                     A93                       A101  …      A121
163                     A93                       A101  …      A124
179                     A93                       A101  …      A121
186                     A92                       A101  …      A124
187                     A93                       A101  …      A123
213                     A93                       A101  …      A121
330                     A93                       A101  …      A124
430                     A93                       A101  …      A121
438                     A93                       A102  …      A123
536                     A92                       A101  …      A122
554                     A92                       A101  …      A122
606                     A93                       A101  …      A121
624                     A93                       A101  …      A124
723                     A92                       A101  …      A121
756                     A93                       A101  …      A121
774                     A93                       A101  …      A124
779                     A92                       A101  …      A123
807                     A93                       A101  …      A121
846                     A93                       A101  …      A123
883                     A92                       A101  …      A123
917                     A93                       A101  …      A124

    Age in years  Other installment plans  Housing  \
0             67                    A143     A152
75            66                    A143     A153
137           66                    A143     A152
163           70                    A141     A153
179           65                    A143     A152
186           74                    A141     A153
187           68                    A143     A153
213           66                    A143     A152
330           75                    A143     A153
430           74                    A143     A152
438           65                    A143     A152
536           75                    A143     A152
554           67                    A143     A152
```

```
606           74                      A143    A152
624           65                      A143    A153
723           66                      A143    A152
756           74                      A143    A152
774           66                      A141    A153
779           67                      A143    A152
807           65                      A143    A152
846           68                      A143    A151
883           65                      A143    A152
917           68                      A141    A152

    Number of existing credits at this bank   Job  \
0                                          2  A173
75                                         2  A174
137                                        1  A172
163                                        1  A174
179                                        2  A173
186                                        1  A174
187                                        3  A171
213                                        1  A174
330                                        2  A174
430                                        1  A172
438                                        2  A171
536                                        1  A174
554                                        2  A174
606                                        1  A174
624                                        2  A173
723                                        1  A172
756                                        3  A171
774                                        3  A171
779                                        1  A173
807                                        4  A173
846                                        2  A173
883                                        2  A171
917                                        1  A174

    Number of people being liable to provide maintenance for  Telephone  \
0                                                          1        A192
75                                                         1        A191
137                                                        1        A191
163                                                        1        A192
179                                                        1        A191
186                                                        2        A192
187                                                        1        A192
213                                                        1        A192
330                                                        1        A192
430                                                        1        A191
438                                                        1        A191
```

```
536                                                 1              A192
554                                                 1              A192
606                                                 1              A192
624                                                 1              A191
723                                                 1              A191
756                                                 2              A191
774                                                 1              A191
779                                                 1              A192
807                                                 1              A191
846                                                 1              A191
883                                                 1              A191
917                                                 1              A192

    foreign worker Receive/ Not receive credit
0             A201                                1
75            A201                                1
137           A201                                2
163           A201                                1
179           A201                                1
186           A201                                2
187           A201                                1
213           A201                                2
330           A201                                1
430           A201                                1
438           A201                                1
536           A201                                1
554           A201                                1
606           A201                                1
624           A201                                2
723           A201                                1
756           A202                                1
774           A201                                1
779           A201                                1
807           A201                                1
846           A201                                2
883           A201                                1
917           A201                                2

[23 rows x 21 columns]
```

[21]:
```python
import pandas as pd

def remove_outliers_iqr(df, columns):
    for column in columns:
        # Calculate Q1 (25th percentile) and Q3 (75th percentile)
        Q1 = df[column].quantile(0.25)
        Q3 = df[column].quantile(0.75)
```

```
        # Calculate the IQR
        IQR = Q3 - Q1
        # Determine outliers
        lower_bound = Q1 - 1.5 * IQR
        upper_bound = Q3 + 1.5 * IQR
        # Filter out the outliers and update the main dataframe
        df = df[(df[column] >= lower_bound) & (df[column] <= upper_bound)]

    return df

# List of numerical columns for which you want to remove outliers
numerical_columns = ['Credit amount', 'Duration in month', 'Age in years']

# Remove outliers from the main DataFrame
df = remove_outliers_iqr(df, numerical_columns)

# Show the shape of the updated DataFrame
print("Updated DataFrame shape:", df.shape)
```

```
Updated DataFrame shape: (862, 21)
```

[22]:
```
df.shape
```

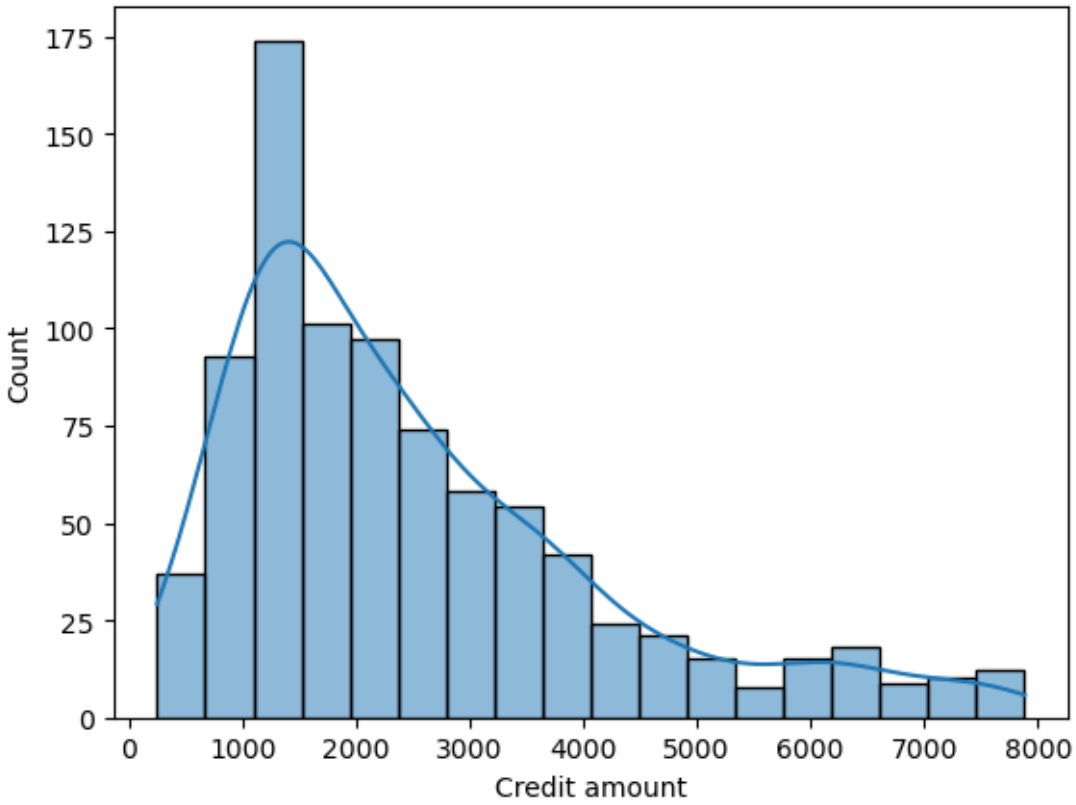
[22]:
```
(862, 21)
```

[23]:
```
df.describe()
```

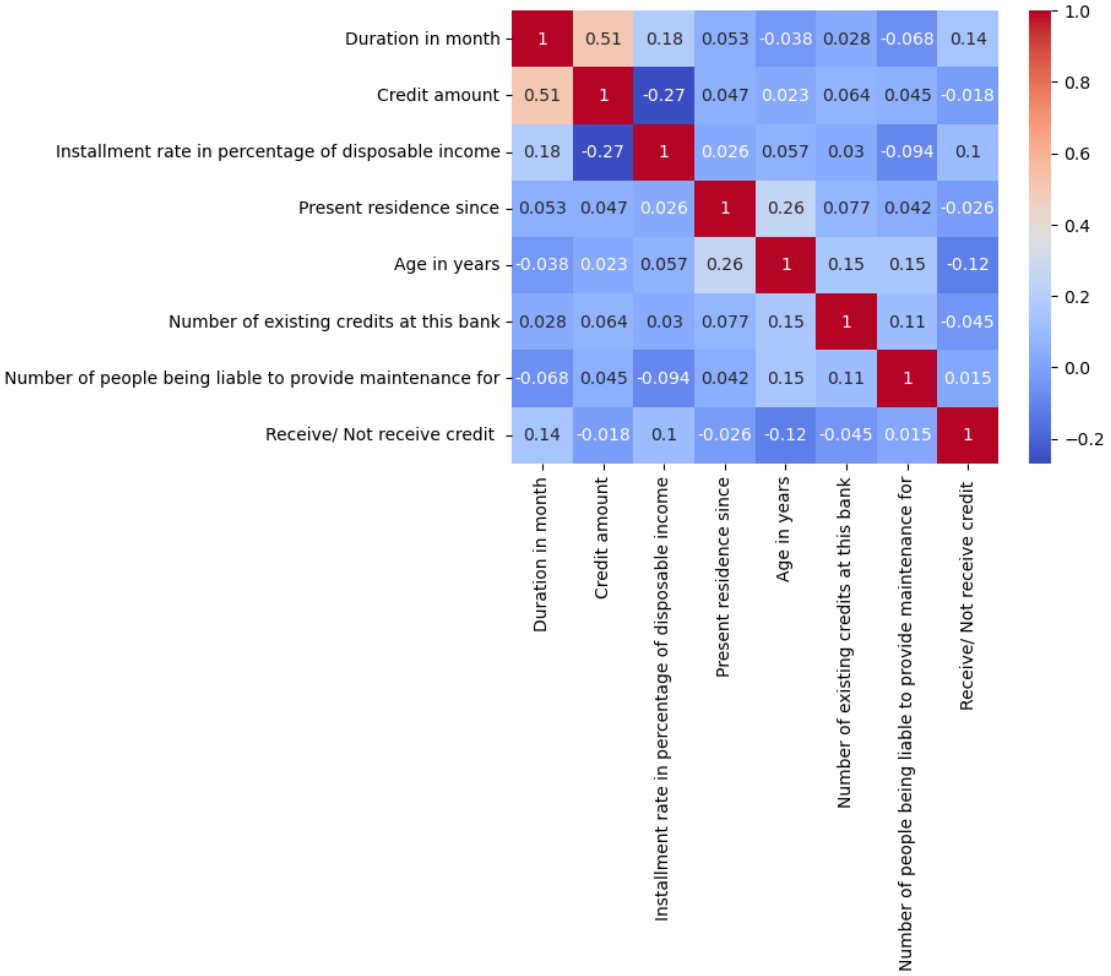
[23]:
```
       Duration in month  Credit amount  \
count         862.000000     862.000000
mean           18.116009    2531.432715
std             8.690069    1663.627810
min             4.000000     250.000000
25%            12.000000    1308.250000
50%            18.000000    2066.500000
75%            24.000000    3355.000000
max            42.000000    7882.000000

       Installment rate in percentage of disposable income  \
count                                          862.000000
mean                                             3.017401
std                                              1.107853
min                                              1.000000
25%                                              2.000000
50%                                              3.000000
75%                                              4.000000
max                                              4.000000
```

```
       Present residence since  Age in years  \
count               862.000000    862.000000
mean                  2.823666     34.698376
std                   1.103893     10.264763
min                   1.000000     19.000000
25%                   2.000000     27.000000
50%                   3.000000     33.000000
75%                   4.000000     41.000000
max                   4.000000     64.000000

       Number of existing credits at this bank  \
count                                862.000000
mean                                   1.397912
std                                    0.564654
min                                    1.000000
25%                                    1.000000
50%                                    1.000000
75%                                    2.000000
max                                    4.000000

       Number of people being liable to provide maintenance for  \
count                                          862.000000
mean                                             1.153132
std                                              0.360324
min                                              1.000000
25%                                              1.000000
50%                                              1.000000
75%                                              1.000000
max                                              2.000000

       Receive/ Not receive credit
count                    862.000000
mean                       1.264501
std                        0.441323
min                        1.000000
25%                        1.000000
50%                        1.000000
75%                        2.000000
max                        2.000000
```

[24]:
```python
import seaborn as sns
import matplotlib.pyplot as plt

# Example: Distribution of Credit Amount
sns.histplot(df['Credit amount'], kde=True)
plt.show()
```



[37]:
```
import numpy as np
# Select only numeric columns for correlation matrix
numeric_df = df.select_dtypes(include=[np.number])

# Calculate the correlation matrix
corr_matrix = numeric_df.corr()

# Plot the heatmap of the correlation matrix
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')
plt.show()
```



[39]:
```
numeric_df = df.select_dtypes(include=[float, int])
```

[41]:
```
numeric_df
```

[41]:

	Duration in month	Credit amount	\
2	12	2096	
3	42	7882	
4	24	4870	
6	24	2835	
7	36	6948	
..	…	…	
993	36	3959	
994	12	2390	
995	12	1736	
996	30	3857	
997	12	804	

```
     Installment rate in percentage of disposable income  \
2                                                      2
3                                                      2
4                                                      3
6                                                      3
7                                                      2
..                                                   …
993                                                    4
994                                                    4
995                                                    3
996                                                    4
997                                                    4

     Present residence since  Age in years  \
2                          3            49
3                          4            45
4                          4            53
6                          4            53
7                          2            35
..                       …            …
993                        3            30
994                        3            50
995                        4            31
996                        4            40
997                        4            38

     Number of existing credits at this bank  \
2                                          1
3                                          1
4                                          2
6                                          1
7                                          1
..                                       …
993                                        1
994                                        1
995                                        1
996                                        1
997                                        1

     Number of people being liable to provide maintenance for  \
2                                                      2
3                                                      2
4                                                      2
6                                                      1
7                                                      1
..                                                   …
```

```
993                                                   1
994                                                   1
995                                                   1
996                                                   1
997                                                   1

     Receive/ Not receive credit
2                               1
3                               1
4                               2
6                               1
7                               1
..                            ...
993                             1
994                             1
995                             1
996                             1
997                             1

[862 rows x 8 columns]
```

[43]:
```
df.dtypes
```

[43]:
```
Status of existing checking account                          object
Duration in month                                             int64
Credit history                                               object
Purpose                                                      object
Credit amount                                                 int64
Savings account/bonds                                        object
Present employment since                                     object
Installment rate in percentage of disposable income           int64
Personal status and sex                                      object
Other debtors / guarantors                                   object
Present residence since                                       int64
Property                                                     object
Age in years                                                  int64
Other installment plans                                      object
Housing                                                      object
Number of existing credits at this bank                       int64
Job                                                          object
Number of people being liable to provide maintenance for      int64
Telephone                                                    object
foreign worker                                               object
Receive/ Not receive credit                                   int64
dtype: object
```

[45]:
```
df.columns = df.columns.str.strip()
```

[47]:
```
df.columns
```

[47]:
```
Index(['Status of existing checking account', 'Duration in month',
       'Credit history', 'Purpose', 'Credit amount', 'Savings account/bonds',
       'Present employment since',
       'Installment rate in percentage of disposable income',
       'Personal status and sex', 'Other debtors / guarantors',
       'Present residence since', 'Property', 'Age in years',
       'Other installment plans', 'Housing',
       'Number of existing credits at this bank', 'Job',
       'Number of people being liable to provide maintenance for', 'Telephone',
       'foreign worker', 'Receive/ Not receive credit'],
      dtype='object')
```

[49]:
```
df['Receive/ Not receive credit'].value_counts()
```

[49]:
```
Receive/ Not receive credit
1    634
2    228
Name: count, dtype: int64
```

[51]:
```
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
df[['Credit amount', 'Duration in month', 'Age in years']] = scaler.
 ↪fit_transform(df[['Credit amount', 'Duration in month', 'Age in years']])
```

[71]:
```
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, accuracy_score

# Define X and y
X = df.drop(columns=['Receive/ Not receive credit'])  # Drop the target column
y = df['Receive/ Not receive credit']  # Define the target variable

# Identify categorical columns
categorical_cols = X.select_dtypes(include=['object', 'category']).columns

# Define the preprocessor
preprocessor = ColumnTransformer(transformers=[('cat',␣
 ↪OneHotEncoder(handle_unknown='ignore'), categorical_cols)],␣
 ↪remainder='passthrough')

# Create a pipeline that first transforms the data and then applies the model
```

```
model_pipeline = Pipeline(steps=[('preprocessor', preprocessor), ('model',␣
 ↪RandomForestClassifier())])

# Split into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,␣
 ↪random_state=42)

# Train the model using the pipeline
model_pipeline.fit(X_train, y_train)

# Make predictions
y_pred = model_pipeline.predict(X_test)

# Evaluate the model
print(classification_report(y_test, y_pred))
print("Accuracy:", accuracy_score(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       0.75      0.30      0.43        70
           1       0.79      0.96      0.87       189

    accuracy                           0.78       259
   macro avg       0.77      0.63      0.65       259
weighted avg       0.78      0.78      0.75       259

Accuracy: 0.7837837837837838
```

[91]:
```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
from sklearn import preprocessing
import statsmodels.api as sm
```

[93]:
```
h11=['Status of existing checking account', 'Duration in month',
       'Credit history', 'Savings account/bonds',
       'Installment rate in percentage of disposable income']
```

[95]:
```
df['Receive/ Not receive credit']=df['Receive/ Not receive credit'].
 ↪replace(to_replace = 2, value =0)
```

[99]:
```
# Assuming h11 is a list of features
categorical_cols = df[h11].select_dtypes(include=['object']).columns.tolist()
```

```
# Define ColumnTransformer to apply one-hot encoding to categorical columns
preprocessor = ColumnTransformer(
    transformers=[
        ('cat', OneHotEncoder(handle_unknown='ignore'), categorical_cols)
    ],
    remainder='passthrough'  # Keep non-categorical columns as they are
)

# Split the data into train and test sets
train, test = train_test_split(df, test_size=0.2)

# Separate features and target variable
train_x = train[h11]
train_y = train['Receive/ Not receive credit']
test_x = test[h11]
test_y = test['Receive/ Not receive credit']

# Apply the transformation to the training and test sets
train_x_encoded = preprocessor.fit_transform(train_x)
test_x_encoded = preprocessor.transform(test_x)

# Scale the encoded features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(train_x_encoded)
X_scaled2 = scaler.transform(test_x_encoded)

# Now X_scaled and X_scaled2 are ready for modeling
```

[101]:
```
X_scaled
```

[101]:
```
array([[-0.59914469,  1.73372834, -0.25478421, …, -0.45691567,
        -0.70609683,  0.88953886],
       [-0.59914469, -0.57679163, -0.25478421, …, -0.45691567,
        -0.70609683,  0.88953886],
       [ 1.66904592, -0.57679163, -0.25478421, …, -0.45691567,
        -0.00508715, -1.83442681],
       …,
       [ 1.66904592, -0.57679163, -0.25478421, …, -0.45691567,
        -1.40710651, -0.92643825],
       [-0.59914469, -0.57679163,  3.92489005, …, -0.45691567,
        -0.70609683, -0.92643825],
       [ 1.66904592, -0.57679163, -0.25478421, …, -0.45691567,
        -1.40710651, -0.01844969]])
```

[103]:
```
X_scaled2
```

[103]:
```
array([[-0.59914469, -0.57679163, -0.25478421, …,  2.18858766,
         0.69592253, -0.01844969],
       [-0.59914469, -0.57679163, -0.25478421, …, -0.45691567,
         0.69592253,  0.88953886],
       [ 1.66904592, -0.57679163, -0.25478421, …, -0.45691567,
        -0.93976673, -1.83442681],
       …,
       [-0.59914469, -0.57679163, -0.25478421, …,  2.18858766,
        -0.70609683,  0.88953886],
       [-0.59914469, -0.57679163,  3.92489005, …, -0.45691567,
         1.39693221,  0.88953886],
       [-0.59914469,  1.73372834, -0.25478421, …, -0.45691567,
        -0.47242694, -1.83442681]])
```

[111]:
```
# Assume df is your DataFrame and 'Receive/ Not receive credit' is the target␣
 ↪column
X = df.drop(columns=['Receive/ Not receive credit'])  # Feature matrix
y = df['Receive/ Not receive credit']  # Target variable

# Identify categorical columns
categorical_cols = X.select_dtypes(include=['object', 'category']).columns

# Define the preprocessor
preprocessor = ColumnTransformer(transformers=[('cat',␣
 ↪OneHotEncoder(handle_unknown='ignore'),␣
 ↪categorical_cols)],remainder='passthrough')

# Transform the features
X_encoded = preprocessor.fit_transform(X)

# Convert X_encoded to a DataFrame
X_encoded_df = pd.DataFrame(X_encoded, columns=preprocessor.
 ↪get_feature_names_out(), index=X.index)

# Split into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X_encoded_df, y,␣
 ↪test_size=0.3, random_state=42, stratify=y)

# Align indices of y_train and X_train
X_train, y_train = X_train.align(y_train, join='inner', axis=0)

# Fit the logistic regression model using statsmodels
logit_model = sm.Logit(y_train, X_train)
result = logit_model.fit()

# Print the model summary
print(result.summary2())
```

```
Warning: Maximum number of iterations has been exceeded.
         Current function value: 0.394781
         Iterations: 35
                                                                  Results: Logit
================================================================================
================================================
Model:                                  Logit
Method:                                 MLE
Dependent Variable:                     Receive/ Not receive credit
Pseudo R-squared:                       0.316
Date:                                   2024-08-12 12:24
AIC:                                    574.1062
No. Observations:                       603
BIC:                                    789.8001
Df Model:                               48
Log-Likelihood:                         -238.05
Df Residuals:                           554
LL-Null:                                -347.85
Converged:                              0.0000
LLR p-value:                            8.6345e-24
No. Iterations:                         35.0000
Scale:                                  1.0000
--------------------------------------------------------------------------------
------------------------------------------------
                                                                          Coef.
Std.Err.      z    P>|z|      [0.025         0.975]
--------------------------------------------------------------------------------
------------------------------------------------
cat__Status of existing checking account_A11                             -0.4559
nan     nan    nan            nan           nan
cat__Status of existing checking account_A12                             -0.2324
nan     nan    nan            nan           nan
cat__Status of existing checking account_A13                              0.6585
nan     nan    nan            nan           nan
cat__Status of existing checking account_A14                              1.5426
nan     nan    nan            nan           nan
cat__Credit history_A30                                                  -0.9859
nan     nan    nan            nan           nan
cat__Credit history_A31                                                  -0.5977
2778974.0688 -0.0000 1.0000  -5446689.6864  5446688.4911
cat__Credit history_A32                                                   0.7855
4971314.5167  0.0000 1.0000  -9743596.6230  9743598.1940
cat__Credit history_A33                                                   0.9532
nan     nan    nan            nan           nan
cat__Credit history_A34                                                   1.3576
4876696.8113  0.0000 1.0000  -9558148.7561  9558151.4713
cat__Purpose_A40                                                         -2.7996
nan     nan    nan            nan           nan
```

```
cat__Purpose_A41                                                    -0.8246
nan     nan    nan            nan           nan
cat__Purpose_A410                                                   18.1838
nan     nan    nan            nan           nan
cat__Purpose_A42                                                    -2.0529
nan     nan    nan            nan           nan
cat__Purpose_A43                                                    -1.6809
nan     nan    nan            nan           nan
cat__Purpose_A44                                                    -2.3515
nan     nan    nan            nan           nan
cat__Purpose_A45                                                    -2.4517
nan     nan    nan            nan           nan
cat__Purpose_A46                                                    -3.2814
nan     nan    nan            nan           nan
cat__Purpose_A48                                                    -0.1858
nan     nan    nan            nan           nan
cat__Purpose_A49                                                    -1.0424
1959727.9105 -0.0000 1.0000  -3840997.1664  3840995.0816
cat__Savings account/bonds_A61                                      -0.2673
6211881.3108 -0.0000 1.0000 -12175063.9127 12175063.3782
cat__Savings account/bonds_A62                                      -0.0834
6120748.1068 -0.0000 1.0000 -11996445.9310 11996445.7643
cat__Savings account/bonds_A63                                      -0.2604
6126425.2263 -0.0000 1.0000 -12007573.0580 12007572.5372
cat__Savings account/bonds_A64                                       1.6418
6259345.6838  0.0000 1.0000 -12268090.4651 12268093.7488
cat__Savings account/bonds_A65                                       0.4820
6230077.2424  0.0000 1.0000 -12210726.5341 12210727.4981
cat__Present employment since_A71                                   -0.4741
nan     nan    nan            nan           nan
cat__Present employment since_A72                                    0.3543
nan     nan    nan            nan           nan
cat__Present employment since_A73                                    0.4299
nan     nan    nan            nan           nan
cat__Present employment since_A74                                    0.7264
nan     nan    nan            nan           nan
cat__Present employment since_A75                                    0.4764
nan     nan    nan            nan           nan
cat__Personal status and sex_A91                                     0.0631
nan     nan    nan            nan           nan
cat__Personal status and sex_A92                                     0.1735
nan     nan    nan            nan           nan
cat__Personal status and sex_A93                                     0.9772
nan     nan    nan            nan           nan
cat__Personal status and sex_A94                                     0.2990
nan     nan    nan            nan           nan
cat__Other debtors / guarantors_A101                                 0.3438
nan     nan    nan            nan           nan
```

```
cat__Other debtors / guarantors_A102                                -0.6596
nan     nan    nan            nan           nan
cat__Other debtors / guarantors_A103                                 1.8287
nan     nan    nan            nan           nan
cat__Property_A121                                                   0.4080
nan     nan    nan            nan           nan
cat__Property_A122                                                   0.4225
nan     nan    nan            nan           nan
cat__Property_A123                                                   0.2460
nan     nan    nan            nan           nan
cat__Property_A124                                                   0.4364
nan     nan    nan            nan           nan
cat__Other installment plans_A141                                    0.2475
5402024.5618  0.0000 1.0000 -10587773.3373 10587773.8323
cat__Other installment plans_A142                                    0.1769
5344232.4520  0.0000 1.0000 -10474502.9540 10474503.3079
cat__Other installment plans_A143                                    1.0884
5340277.8071  0.0000 1.0000 -10466751.0810 10466753.2578
cat__Housing_A151                                                    0.5866
nan     nan    nan            nan           nan
cat__Housing_A152                                                    0.6429
nan     nan    nan            nan           nan
cat__Housing_A153                                                    0.2833
nan     nan    nan            nan           nan
cat__Job_A171                                                        1.1591
4401592.2282  0.0000 1.0000  -8626961.0827  8626963.4009
cat__Job_A172                                                        0.3590
4408650.8094  0.0000 1.0000  -8640796.4477  8640797.1658
cat__Job_A173                                                        0.0511
4407581.2620  0.0000 1.0000  -8638700.4813  8638700.5834
cat__Job_A174                                                       -0.0564
4393897.1590 -0.0000 1.0000  -8611880.2397  8611880.1270
cat__Telephone_A191                                                  0.4789
nan     nan    nan            nan           nan
cat__Telephone_A192                                                  1.0339
nan     nan    nan            nan           nan
cat__foreign worker_A201                                             0.0602
nan     nan    nan            nan           nan
cat__foreign worker_A202                                             1.4526
nan     nan    nan            nan           nan
remainder__Duration in month                                        -0.4087
0.1496 -2.7316 0.0063        -0.7019       -0.1154
remainder__Credit amount                                             0.1629
0.1682  0.9686 0.3327        -0.1668        0.4926
remainder__Installment rate in percentage of disposable income      -0.2929
0.1297 -2.2582 0.0239        -0.5471       -0.0387
remainder__Present residence since                                   0.0545
0.1208  0.4509 0.6521        -0.1823        0.2913
```

```
remainder__Age in years                                                    0.0687
0.1495  0.4597 0.6458        -0.2243        0.3618
remainder__Number of existing credits at this bank                        -0.2350
0.2859 -0.8221 0.4110        -0.7953        0.3253
remainder__Number of people being liable to provide maintenance for -0.7627
0.3562 -2.1410 0.0323        -1.4609       -0.0645
==============================================================================
==============================================
```

```
C:\Users\HP\.anaconda\anaconda3\Lib\site-packages\statsmodels\base\model.py:607:
ConvergenceWarning: Maximum Likelihood optimization failed to converge. Check
mle_retvals
  warnings.warn("Maximum Likelihood optimization failed to "
```

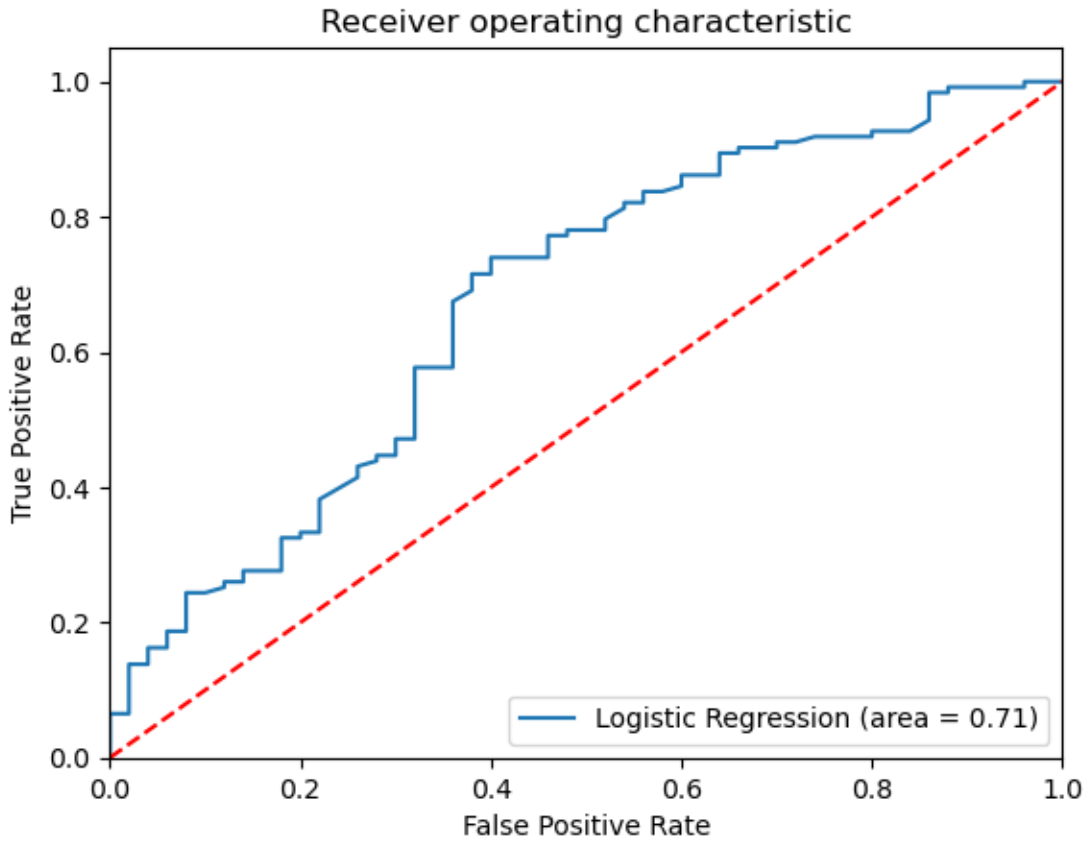
[107]:
```
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, confusion_matrix,␣
 ↪classification_report
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

# Assuming h11 is the list of features including categorical and numerical␣
 ↪columns
categorical_cols = df[h11].select_dtypes(include=['object']).columns.tolist()

# Define a ColumnTransformer to apply OneHotEncoder to categorical columns
preprocessor = ColumnTransformer(
    transformers=[
        ('cat', OneHotEncoder(handle_unknown='ignore'), categorical_cols)
    ],
    remainder='passthrough'  # Keep other columns unchanged
)

# Apply the transformation to both train and test data
X_encoded = preprocessor.fit_transform(df[h11])

# Split the data into training and testing sets
train_x, test_x, train_y, test_y = train_test_split(X_encoded, df['Receive/ Not␣
 ↪receive credit'], test_size=0.2, random_state=42)

# Scale the features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(train_x)
X_scaled2 = scaler.transform(test_x)
model = LogisticRegression(penalty='l2', C=0.1, class_weight= "balanced")
```

```
# fit the model with the training data

model.fit(X_scaled,train_y)

## coefficeints of the trained model
print('Coefficient of model :', model.coef_)
#
## intercept of the model
print('Intercept of model',model.intercept_)

# predict the target on the train dataset
predict_train = model.predict(train_x)
#print('Target on train data',predict_train)

# Accuray Score on train dataset
accuracy_train = accuracy_score(train_y,predict_train)
print('accuracy_score on train dataset : ', accuracy_train)

# predict the target on the test dataset
predict_test = model.predict(X_scaled2)
#print('Target on test data',predict_test)

# Accuracy Score on test dataset
accuracy_test = accuracy_score(test_y,predict_test)
print('accuracy_score on test dataset : ', accuracy_test)

from sklearn.metrics import confusion_matrix
confusion_matrix = confusion_matrix(test_y,predict_test)
print(confusion_matrix)

from sklearn.metrics import classification_report
print(classification_report(test_y,predict_test))
```

```
Coefficient of model : [[-0.41220492 -0.22100446  0.00084696  0.57245104
-0.1156956  -0.25160519
  -0.03092211 -0.07224988  0.23257295 -0.16874965 -0.01080701 -0.02789966
   0.14443648  0.15576632 -0.3557308  -0.24389266]]
Intercept of model [0.36104155]
accuracy_score on train dataset :  0.5297532656023222
accuracy_score on test dataset :  0.6763005780346821
[[31 19]
 [37 86]]
              precision    recall  f1-score   support

           0       0.46      0.62      0.53        50
           1       0.82      0.70      0.75       123
```

```
              accuracy                           0.68       173
             macro avg       0.64      0.66      0.64       173
          weighted avg       0.71      0.68      0.69       173
```

[109]:
```
from sklearn.metrics import roc_auc_score
from sklearn.metrics import roc_curve
logit_roc_auc = roc_auc_score(test_y, model.predict(X_scaled2))
fpr, tpr, thresholds = roc_curve(test_y, model.predict_proba(X_scaled2)[:,1])
plt.figure()
plt.plot(fpr, tpr, label='Logistic Regression (area = 0.71)' % logit_roc_auc)
plt.plot([0, 1], [0, 1],'r--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver operating characteristic')
plt.legend(loc="lower right")
plt.savefig('Log_ROC')
plt.show()
```



[113]:
```
from sklearn.model_selection import GridSearchCV
param_grid = {'n_estimators': [100, 200], 'max_depth': [10, 20]}
grid_search = GridSearchCV(estimator=RandomForestClassifier(),␣
 ↪param_grid=param_grid)
grid_search.fit(X_train, y_train)
```

[113]:
```
GridSearchCV(estimator=RandomForestClassifier(),
             param_grid={'max_depth': [10, 20], 'n_estimators': [100, 200]})
```

[141]:
```
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, accuracy_score

# Example DataFrame and target column
df = pd.DataFrame({
    'Feature1': ['A', 'B', 'C', 'A', 'B', 'C'],
    'Feature2': [1, 2, 3, 4, 5, 6],
    'Receive/Not receive credit': [1, 0, 1, 0, 1, 0]
})

# Features and target
X = df.drop(columns=['Receive/Not receive credit'])
y = df['Receive/Not receive credit']

# Identify categorical columns
categorical_cols = X.select_dtypes(include=['object', 'category']).columns

# Define the preprocessor
preprocessor = ColumnTransformer(
    transformers=[('cat', OneHotEncoder(handle_unknown='ignore'),␣
 ↪categorical_cols)],
    remainder='passthrough'
)

# Transform the features
X_encoded = preprocessor.fit_transform(X)
X_encoded_df = pd.DataFrame(X_encoded, columns=preprocessor.
 ↪get_feature_names_out(), index=X.index)

# Split into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X_encoded_df, y,␣
 ↪test_size=0.2, random_state=42, stratify=y)

# Ensure n_neighbors is less than or equal to the number of training samples
knn_model = KNeighborsClassifier(n_neighbors=min(5, len(X_train)))
```

```python
knn_model.fit(X_train, y_train)

# Predict and evaluate
y_pred_knn = knn_model.predict(X_test)
print("KNN Classification Report:")
print(classification_report(y_test, y_pred_knn))
print("KNN Accuracy:", accuracy_score(y_test, y_pred_knn))
```

```
KNN Classification Report:
              precision    recall  f1-score   support

           0       0.50      1.00      0.67         1
           1       0.00      0.00      0.00         1

    accuracy                           0.50         2
   macro avg       0.25      0.50      0.33         2
weighted avg       0.25      0.50      0.33         2

KNN Accuracy: 0.5
```

```
C:\Users\HP\.anaconda\anaconda3\Lib\site-
packages\sklearn\metrics\_classification.py:1509: UndefinedMetricWarning:
Precision is ill-defined and being set to 0.0 in labels with no predicted
samples. Use `zero_division` parameter to control this behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
C:\Users\HP\.anaconda\anaconda3\Lib\site-
packages\sklearn\metrics\_classification.py:1509: UndefinedMetricWarning:
Precision is ill-defined and being set to 0.0 in labels with no predicted
samples. Use `zero_division` parameter to control this behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
C:\Users\HP\.anaconda\anaconda3\Lib\site-
packages\sklearn\metrics\_classification.py:1509: UndefinedMetricWarning:
Precision is ill-defined and being set to 0.0 in labels with no predicted
samples. Use `zero_division` parameter to control this behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
```