

Ex04_Implementation_of_Logistic_Regression_Model_to_Predict_the_Placement_Status

May 8, 2023

```
[38]: import pandas as pd

data = pd.read_csv("Placement_Data.csv")

data.head()
```

```
[38]:   sl_no  gender  ssc_p  ssc_b  hsc_p  hsc_b  hsc_s  degree_p  \
0      1      M  67.00  Others  91.00  Others  Commerce    58.00
1      2      M  79.33  Central  78.33  Others  Science    77.48
2      3      M  65.00  Central  68.00  Central    Arts    64.00
3      4      M  56.00  Central  52.00  Central  Science    52.00
4      5      M  85.80  Central  73.60  Central  Commerce    73.30

      degree_t  workex  etest_p  specialisation  mba_p  status  salary
0  Sci&Tech      No    55.0      Mkt&HR    58.80  Placed  270000.0
1  Sci&Tech     Yes    86.5      Mkt&Fin    66.28  Placed  200000.0
2  Comm&Mgmt      No    75.0      Mkt&Fin    57.80  Placed  250000.0
3  Sci&Tech      No    66.0      Mkt&HR    59.43  Not Placed      NaN
4  Comm&Mgmt      No    96.8      Mkt&Fin    55.50  Placed  425000.0
```

```
[39]: data1 = data.copy()
data1= data1.drop(["sl_no","salary"],axis=1) #feature selection
print(data1.head())
```

```
   gender  ssc_p  ssc_b  hsc_p  hsc_b  hsc_s  degree_p  degree_t  \
0      M  67.00  Others  91.00  Others  Commerce    58.00  Sci&Tech
1      M  79.33  Central  78.33  Others  Science    77.48  Sci&Tech
2      M  65.00  Central  68.00  Central    Arts    64.00  Comm&Mgmt
3      M  56.00  Central  52.00  Central  Science    52.00  Sci&Tech
4      M  85.80  Central  73.60  Central  Commerce    73.30  Comm&Mgmt

   workex  etest_p  specialisation  mba_p  status
0     No    55.0      Mkt&HR    58.80  Placed
1    Yes    86.5      Mkt&Fin    66.28  Placed
2     No    75.0      Mkt&Fin    57.80  Placed
3     No    66.0      Mkt&HR    59.43  Not Placed
```

| | | | | | |
|---|----|------|---------|-------|--------|
| 4 | No | 96.8 | Mkt&Fin | 55.50 | Placed |
|---|----|------|---------|-------|--------|

```
[40]: data1.isnull().sum()
```

```
[40]: gender          0
      ssc_p           0
      ssc_b           0
      hsc_p           0
      hsc_b           0
      hsc_s           0
      degree_p        0
      degree_t        0
      workex          0
      etest_p         0
      specialisation  0
      mba_p           0
      status          0
      dtype: int64
```

```
[41]: data1.duplicated().sum()
```

```
[41]: 0
```

```
[42]: from sklearn.preprocessing import LabelEncoder
      lc = LabelEncoder() #changing values

      data1["gender"] = lc.fit_transform(data1["gender"])
      data1["ssc_b"] = lc.fit_transform(data1["ssc_b"])
      data1["hsc_b"] = lc.fit_transform(data1["hsc_b"])
      data1["hsc_s"] = lc.fit_transform(data1["hsc_s"])
      data1["degree_t"] = lc.fit_transform(data1["degree_t"])
      data1["workex"] = lc.fit_transform(data1["workex"])
      data1["specialisation"] = lc.fit_transform(data1["specialisation"])
      data1["status"] = lc.fit_transform(data1["status"])

      print(data1)
```

| | gender | ssc_p | ssc_b | hsc_p | hsc_b | hsc_s | degree_p | degree_t | workex | \ |
|-----|--------|-------|-------|-------|-------|-------|----------|----------|--------|---|
| 0 | 1 | 67.00 | 1 | 91.00 | 1 | 1 | 58.00 | 2 | 0 | |
| 1 | 1 | 79.33 | 0 | 78.33 | 1 | 2 | 77.48 | 2 | 1 | |
| 2 | 1 | 65.00 | 0 | 68.00 | 0 | 0 | 64.00 | 0 | 0 | |
| 3 | 1 | 56.00 | 0 | 52.00 | 0 | 2 | 52.00 | 2 | 0 | |
| 4 | 1 | 85.80 | 0 | 73.60 | 0 | 1 | 73.30 | 0 | 0 | |
| .. | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 210 | 1 | 80.60 | 1 | 82.00 | 1 | 1 | 77.60 | 0 | 0 | |
| 211 | 1 | 58.00 | 1 | 60.00 | 1 | 2 | 72.00 | 2 | 0 | |
| 212 | 1 | 67.00 | 1 | 67.00 | 1 | 1 | 73.00 | 0 | 1 | |
| 213 | 0 | 74.00 | 1 | 66.00 | 1 | 1 | 58.00 | 0 | 0 | |

| | | | | | | | | | |
|-----|---|-------|---|-------|---|---|-------|---|---|
| 214 | 1 | 62.00 | 0 | 58.00 | 1 | 2 | 53.00 | 0 | 0 |
|-----|---|-------|---|-------|---|---|-------|---|---|

| | etest_p | specialisation | mba_p | status |
|-----|---------|----------------|---------|--------|
| 0 | 55.0 | | 1 58.80 | 1 |
| 1 | 86.5 | | 0 66.28 | 1 |
| 2 | 75.0 | | 0 57.80 | 1 |
| 3 | 66.0 | | 1 59.43 | 0 |
| 4 | 96.8 | | 0 55.50 | 1 |
| .. | ... | ... | ... | ... |
| 210 | 91.0 | | 0 74.49 | 1 |
| 211 | 74.0 | | 0 53.62 | 1 |
| 212 | 59.0 | | 0 69.72 | 1 |
| 213 | 70.0 | | 1 60.23 | 1 |
| 214 | 89.0 | | 1 60.22 | 0 |

[215 rows x 13 columns]

```
[43]: y = data1["status"]
      print(y)
```

| | |
|-----|----|
| 0 | 1 |
| 1 | 1 |
| 2 | 1 |
| 3 | 0 |
| 4 | 1 |
| .. | .. |
| 210 | 1 |
| 211 | 1 |
| 212 | 1 |
| 213 | 1 |
| 214 | 0 |

Name: status, Length: 215, dtype: int64

```
[44]: x = data1.iloc[:, :-1]
      print(x)
```

| | gender | ssc_p | ssc_b | hsc_p | hsc_b | hsc_s | degree_p | degree_t | workex | \ |
|-----|--------|-------|-------|-------|-------|-------|----------|----------|--------|-----|
| 0 | 1 | 67.00 | 1 | 91.00 | 1 | 1 | 58.00 | 2 | 0 | |
| 1 | 1 | 79.33 | 0 | 78.33 | 1 | 2 | 77.48 | 2 | 1 | |
| 2 | 1 | 65.00 | 0 | 68.00 | 0 | 0 | 64.00 | 0 | 0 | |
| 3 | 1 | 56.00 | 0 | 52.00 | 0 | 2 | 52.00 | 2 | 0 | |
| 4 | 1 | 85.80 | 0 | 73.60 | 0 | 1 | 73.30 | 0 | 0 | |
| .. | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 210 | 1 | 80.60 | 1 | 82.00 | 1 | 1 | 77.60 | 0 | 0 | |
| 211 | 1 | 58.00 | 1 | 60.00 | 1 | 2 | 72.00 | 2 | 0 | |
| 212 | 1 | 67.00 | 1 | 67.00 | 1 | 1 | 73.00 | 0 | 1 | |
| 213 | 0 | 74.00 | 1 | 66.00 | 1 | 1 | 58.00 | 0 | 0 | |
| 214 | 1 | 62.00 | 0 | 58.00 | 1 | 2 | 53.00 | 0 | 0 | |

| | etest_p | specialisation | mba_p |
|-----|---------|----------------|-------|
| 0 | 55.0 | 1 | 58.80 |
| 1 | 86.5 | 0 | 66.28 |
| 2 | 75.0 | 0 | 57.80 |
| 3 | 66.0 | 1 | 59.43 |
| 4 | 96.8 | 0 | 55.50 |
| .. | ... | ... | ... |
| 210 | 91.0 | 0 | 74.49 |
| 211 | 74.0 | 0 | 53.62 |
| 212 | 59.0 | 0 | 69.72 |
| 213 | 70.0 | 1 | 60.23 |
| 214 | 89.0 | 1 | 60.22 |

[215 rows x 12 columns]

```
[45]: from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x,y,test_size=0.
↪2,random_state=0)

from sklearn.linear_model import LogisticRegression
lr = LogisticRegression(solver="liblinear")
print(lr.fit(x_train,y_train))
```

LogisticRegression(solver='liblinear')

```
[46]: y_pred = lr.predict(x_test)
print(y_pred)
```

```
[0 1 1 0 1 0 1 1 1 1 1 1 1 0 0 1 0 0 1 0 1 1 1 0 0 1 1 1 1 1 1 1 1 0 0 1 1
 1 1 1 0 0 1]
```

```
[47]: from sklearn.metrics import accuracy_score
accuracy = accuracy_score(y_test,y_pred)
print(accuracy)
```

0.813953488372093

```
[48]: from sklearn.metrics import confusion_matrix
confusion = confusion_matrix(y_test,y_pred)
print(confusion)
```

```
[[11  5]
 [ 3 24]]
```

```
[49]: from sklearn.metrics import classification_report
classification_report1 = classification_report(y_test,y_pred)
print(classification_report1)
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.79 | 0.69 | 0.73 | 16 |
| 1 | 0.83 | 0.89 | 0.86 | 27 |
| accuracy | | | 0.81 | 43 |
| macro avg | 0.81 | 0.79 | 0.80 | 43 |
| weighted avg | 0.81 | 0.81 | 0.81 | 43 |

```
[50]: #for prediction lets take the first value from data 1
```

```
prediction = [1,67,1,91,1,1,58,2,0,55,1,58.80]
print(lr.predict([prediction])) # status should be 1
```

```
[1]
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning: X does
not have valid feature names, but LogisticRegression was fitted with feature
names
  warnings.warn(
```

```
[51]: #now we predict for random value asuuming gender ssc_p ssc_b .... be
```

```
prediction = [1,80,1,90,1,1,90,1,0,85,1,85]
print(lr.predict([prediction]))
```

```
[0]
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning: X does
not have valid feature names, but LogisticRegression was fitted with feature
names
  warnings.warn(
```