

# Data Analytics II

May 2, 2022

```
[1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
[2]: data = pd.read_csv("Social_Network_Ads.csv")
```

```
[3]: data.head()
```

```
[3]:
```

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0

```
[4]: from sklearn.preprocessing import LabelEncoder
lb = LabelEncoder()
data['Gender'] = pd.DataFrame(lb.fit_transform(data['Gender']))
```

```
[5]: data.head()
```

```
[5]:
```

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	1	19	19000	0
1	15810944	1	35	20000	0
2	15668575	0	26	43000	0
3	15603246	0	27	57000	0
4	15804002	1	19	76000	0

```
[6]: data.isnull().sum()
```

```
[6]: User ID      0
Gender        0
Age           0
EstimatedSalary  0
Purchased     0
dtype: int64
```

```
[7]: x = data.drop(['Purchased'], axis=1)
     y = data['Purchased']
```

```
[8]: from sklearn.model_selection import train_test_split
     xtrain, xtest, ytrain, ytest = train_test_split(x, y, test_size = 0.8,
     ↪random_state=0)
```

```
[9]: from sklearn.linear_model import LogisticRegression
     lr = LogisticRegression()
     model = lr.fit(xtrain,ytrain)
```

```
[10]: ytest_pred = lr.predict(xtest)
```

```
[11]: df = pd.DataFrame(ytest_pred, ytest)
```

```
[12]: from sklearn.metrics import
     ↪precision_score,confusion_matrix,accuracy_score,recall_score
     cm = confusion_matrix(ytest,ytest_pred)
     print(cm)
```

```
[[193   8]
 [ 68  51]]
```

```
[13]: ps = precision_score(ytest,ytest_pred)
     print(ps)
```

```
0.864406779661017
```

```
[14]: accs = accuracy_score(ytest,ytest_pred)
     print(accs)
```

```
0.7625
```

```
[15]: rs = recall_score(ytest,ytest_pred)
     print(rs)
```

```
0.42857142857142855
```