ex3

July 12, 2024

[ ]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import cross_val_score
```

[ ]:
```
df = pd.read_csv('datasets\churn-bigml-20.csv')
df.head()
```

[ ]:
```
  State  Account length  Area code International plan Voice mail plan  \
0    LA             117        408                 No              No
1    IN              65        415                 No              No
2    NY             161        415                 No              No
3    SC             111        415                 No              No
4    HI              49        510                 No              No

   Number vmail messages  Total day minutes  Total day calls  \
0                      0              184.5               97
1                      0              129.1              137
2                      0              332.9               67
3                      0              110.4              103
4                      0              119.3              117

   Total day charge  Total eve minutes  Total eve calls  Total eve charge  \
0             31.37              351.6               80             29.89
1             21.95              228.5               83             19.42
2             56.59              317.8               97             27.01
3             18.77              137.3              102             11.67
4             20.28              215.1              109             18.28

   Total night minutes  Total night calls  Total night charge  \
0                215.8                 90                9.71
1                208.8                111                9.40
2                160.6                128                7.23
3                189.6                105                8.53
4                178.7                 90                8.04

   Total intl minutes  Total intl calls  Total intl charge  \
```

```
0                 8.7                4              2.35
1                12.7                6              3.43
2                 5.4                9              1.46
3                 7.7                6              2.08
4                11.1                1              3.00

   Customer service calls  Churn
0                       1  False
1                       4   True
2                       4   True
3                       2  False
4                       1  False
```

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

[ ]:
```
(667, 20)
```

[ ]:
```
import seaborn as sns

#sns.pairplot(df)
```

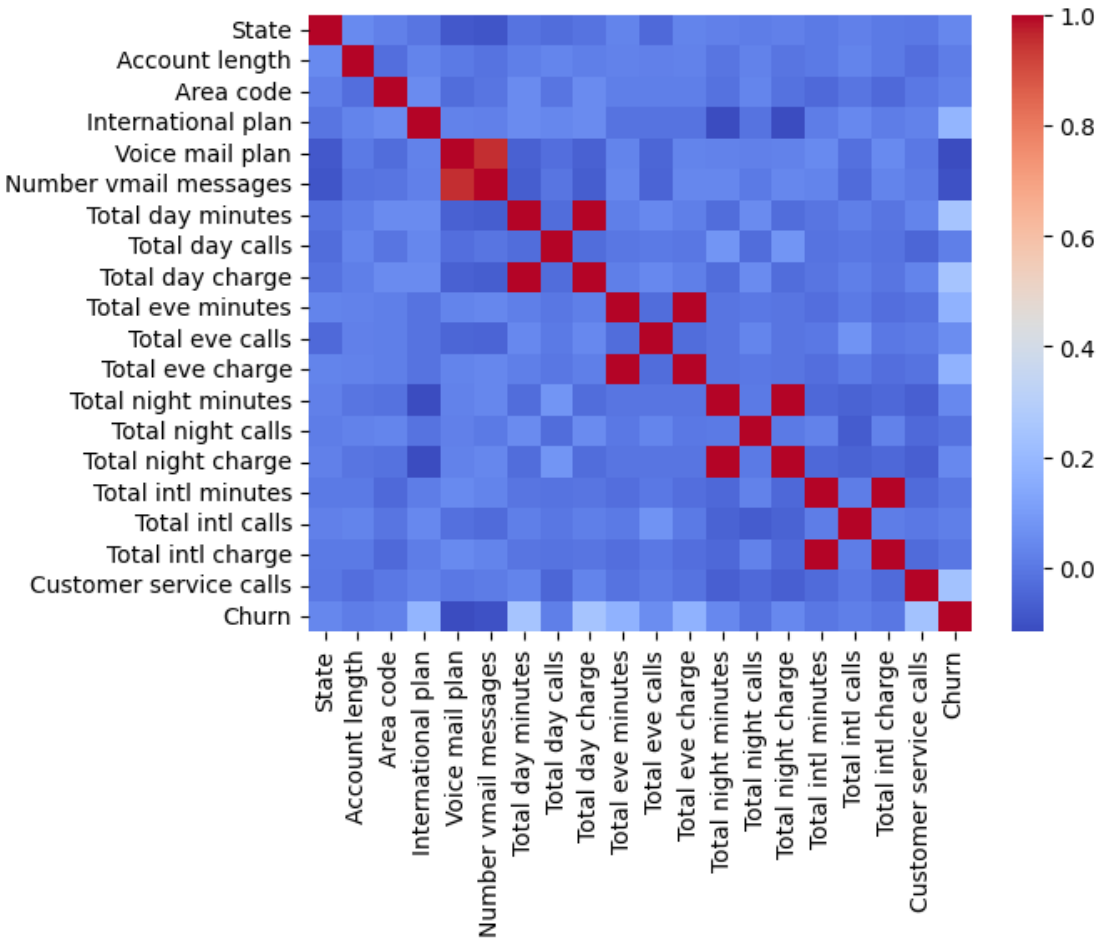
[ ]:
```
model = LogisticRegression(max_iter=1580)
le = LabelEncoder()

df['International plan'] = le.fit_transform(df['International plan'])
df['Voice mail plan'] = le.fit_transform(df['Voice mail plan'])
df['Churn'] = le.fit_transform(df['Churn'])
df['State'] = le.fit_transform(df['State'])

X = df.drop('Churn', axis=1)
y = df['Churn']
```

[ ]:
```
sns.heatmap(df.corr(), cmap='coolwarm')
```

[ ]:
```
<Axes: >
```



[ ]:
```
#After Pre-processing

X.head()
```

[ ]:

	State	Account length	Area code	International plan	Voice mail plan	\
0	18	117	408	0	0	
1	15	65	415	0	0	
2	34	161	415	0	0	
3	40	111	415	0	0	
4	11	49	510	0	0	

	Number vmail messages	Total day minutes	Total day calls	\
0	0	184.5	97	
1	0	129.1	137	
2	0	332.9	67	
3	0	110.4	103	
4	0	119.3	117	

```
   Total day charge  Total eve minutes  Total eve calls  Total eve charge  \
0             31.37              351.6               80             29.89
1             21.95              228.5               83             19.42
2             56.59              317.8               97             27.01
3             18.77              137.3              102             11.67
4             20.28              215.1              109             18.28

   Total night minutes  Total night calls  Total night charge  \
0                215.8                 90                9.71
1                208.8                111                9.40
2                160.6                128                7.23
3                189.6                105                8.53
4                178.7                 90                8.04

   Total intl minutes  Total intl calls  Total intl charge  \
0                 8.7                 4               2.35
1                12.7                 6               3.43
2                 5.4                 9               1.46
3                 7.7                 6               2.08
4                11.1                 1               3.00

   Customer service calls
0                       1
1                       4
2                       4
3                       2
4                       1
```

[ ]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,␣
↪random_state=42)

model.fit(X_train, y_train)
print(f"Accuracy :{model.score(X_test, y_test)}")
```

```
Accuracy :0.8905472636815921
```

[ ]:
```
y_pred = model.predict(X_test)
y_pred
```

[ ]:
```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
       0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0,
       0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
       0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
```

```
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
       0, 0, 0], dtype=int64)
```

[ ]:
```
import numpy as np

individual_values = np.array([18,117,408,0,0,0,184.5,97,31.37,351.6,80,29.
 ↪89,215.8,90,9.71,8.7,4,2.35,1]).reshape(1,- 1)  # Select the rows you want␣
 ↪to predict
predictions = model.predict(individual_values)
print(predictions)
```

```
[0]

c:\Python311\Lib\site-packages\sklearn\base.py:439: UserWarning: X does not have
valid feature names, but LogisticRegression was fitted with feature names
  warnings.warn(
```