Aula10

October 14, 2024

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

[3]:
```
df = pd.read_csv("titanic.csv")
df.head()
```

[3]:
```
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3

                                                Name     Sex   Age  SibSp  \
0                            Braund, Mr. Owen Harris    male  22.0      1
1  Cumings, Mrs. John Bradley (Florence Briggs Th…  female  38.0      1
2                             Heikkinen, Miss. Laina  female  26.0      0
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                           Allen, Mr. William Henry    male  35.0      0

   Parch            Ticket     Fare Cabin Embarked
0      0         A/5 21171   7.2500   NaN        S
1      0          PC 17599  71.2833   C85        C
2      0  STON/O2. 3101282   7.9250   NaN        S
3      0            113803  53.1000  C123        S
4      0            373450   8.0500   NaN        S
```

1 Dados Numericos

[16]:
```
dfNumerico = df[["PassengerId","Age","Fare"]]
dfNumerico = dfNumerico.dropna()
dfNumerico.head()
```

[16]:

```
   PassengerId   Age     Fare
0            1  22.0   7.2500
1            2  38.0  71.2833
2            3  26.0   7.9250
3            4  35.0  53.1000
4            5  35.0   8.0500
```

1.1 Importando pre processadores

[14]:

```
from sklearn.preprocessing import MaxAbsScaler, MinMaxScaler, StandardScaler,␣
 ↪Normalizer, RobustScaler
```

[14]:

```
   PassengerId     Age      Fare
0     0.001122  0.2750  0.014151
1     0.002245  0.4750  0.139136
2     0.003367  0.3250  0.015469
3     0.004489  0.4375  0.103644
4     0.005612  0.4375  0.015713
```

1.2 Max abs Scaler

[15]:

```
max_abs_scaler = MaxAbsScaler()
dfMaxAbsScaler = pd.DataFrame(max_abs_scaler.fit_transform(dfNumerico),␣
 ↪columns=dfNumerico.columns)
dfMaxAbsScaler.head()
```

[15]:

```
   PassengerId     Age      Fare
0     0.001122  0.2750  0.014151
1     0.002245  0.4750  0.139136
2     0.003367  0.3250  0.015469
3     0.004489  0.4375  0.103644
4     0.005612  0.4375  0.015713
```

1.3 MinMaxScaler

[17]:

```
min_max_scaler = MinMaxScaler()
df_min_max_scaled = pd.DataFrame(min_max_scaler.fit_transform(dfNumerico),␣
 ↪columns=dfNumerico.columns)
df_min_max_scaled.head()
```

[17]:

```
   PassengerId       Age      Fare
0     0.000000  0.271174  0.014151
1     0.001124  0.472229  0.139136
2     0.002247  0.321438  0.015469
3     0.003371  0.434531  0.103644
4     0.004494  0.434531  0.015713
```

1.4 standard_scaler

```
[18]: standard_scaler = StandardScaler()
      df_standard_scaled = pd.DataFrame(standard_scaler.fit_transform(dfNumerico),␣
       ↪columns=dfNumerico.columns)
      df_standard_scaled.head()
```

```
[18]:    PassengerId       Age      Fare
      0    -1.728532 -0.530377 -0.518978
      1    -1.724670  0.571831  0.691897
      2    -1.720808 -0.254825 -0.506214
      3    -1.716946  0.365167  0.348049
      4    -1.713084  0.365167 -0.503850
```

1.5 Normalizer

```
[19]: normalizer = Normalizer()
      df_normalized = pd.DataFrame(normalizer.fit_transform(dfNumerico),␣
       ↪columns=dfNumerico.columns)
      df_normalized.head()
```

```
[19]:    PassengerId       Age      Fare
      0     0.043131  0.948873  0.312697
      1     0.024751  0.470273  0.882174
      2     0.109705  0.950778  0.289804
      3     0.062772  0.549253  0.833295
      4     0.137892  0.965245  0.222006
```

1.6 RobustScaler

```
[20]: robust_scaler = RobustScaler()
      df_robust_scaled = pd.DataFrame(robust_scaler.fit_transform(dfNumerico),␣
       ↪columns=dfNumerico.columns)
      df_robust_scaled.head()
```

```
[20]:    PassengerId       Age      Fare
      0    -0.974753 -0.335664 -0.335309
      1    -0.972558  0.559441  2.193153
      2    -0.970362 -0.111888 -0.308655
      3    -0.968167  0.391608  1.475155
      4    -0.965971  0.391608 -0.303720
```

2 Dados Categoricos

[9]:

```
dfCategorico = df[["Survived", "Pclass", "Sex", "SibSp", "Parch",␣
 ↪"Embarked","Cabin","Ticket"]]
dfCategorico.head()
```

[9]:

```
   Survived  Pclass     Sex  SibSp  Parch Embarked Cabin            Ticket
0         0       3    male      1      0        S   NaN         A/5 21171
1         1       1  female      1      0        C   C85          PC 17599
2         1       3  female      0      0        S   NaN  STON/O2. 3101282
3         1       1  female      1      0        S  C123            113803
4         0       3    male      0      0        S   NaN            373450
```

2.1 importando pre processadores

[34]:

```
from sklearn.preprocessing import LabelEncoder, OneHotEncoder
columnsToEncode = ['Survived', 'Pclass', 'Sex', 'SibSp', 'Parch', 'Embarked']
```

2.2 OneHotEncoder

[42]:

```
encoder = OneHotEncoder()
encodedData = encoder.fit_transform(df[columnsToEncode])
encodedDf = pd.DataFrame(encodedData.toarray(), columns=encoder.
 ↪get_feature_names_out(columnsToEncode))
encodedDf.head()
```

[42]:

```
   Survived_0  Survived_1  Pclass_1  Pclass_2  Pclass_3  Sex_female  Sex_male  \
0         1.0         0.0       0.0       0.0       1.0         0.0       1.0
1         0.0         1.0       1.0       0.0       0.0         1.0       0.0
2         0.0         1.0       0.0       0.0       1.0         1.0       0.0
3         0.0         1.0       1.0       0.0       0.0         1.0       0.0
4         1.0         0.0       0.0       0.0       1.0         0.0       1.0

   SibSp_0  SibSp_1  SibSp_2  …  Parch_1  Parch_2  Parch_3  Parch_4  \
0      0.0      1.0      0.0  …      0.0      0.0      0.0      0.0
1      0.0      1.0      0.0  …      0.0      0.0      0.0      0.0
2      1.0      0.0      0.0  …      0.0      0.0      0.0      0.0
3      0.0      1.0      0.0  …      0.0      0.0      0.0      0.0
4      1.0      0.0      0.0  …      0.0      0.0      0.0      0.0

   Parch_5  Parch_6  Embarked_C  Embarked_Q  Embarked_S  Embarked_nan
0      0.0      0.0         0.0         0.0         1.0           0.0
1      0.0      0.0         1.0         0.0         0.0           0.0
2      0.0      0.0         0.0         0.0         1.0           0.0
3      0.0      0.0         0.0         0.0         1.0           0.0
4      0.0      0.0         0.0         0.0         1.0           0.0
```

```
[5 rows x 25 columns]
```

2.3 labelEncoder

[43]:
```
labelEncoder = LabelEncoder()
```

[44]:
```
df[columnsToEncode] = df[columnsToEncode].apply(lambda col: labelEncoder.
 ↪fit_transform(col))
df.head()
```

[44]:
```
   PassengerId  Survived  Pclass  \
0            1         0       2
1            2         1       0
2            3         1       2
3            4         1       0
4            5         0       2

                                                Name  Sex   Age  SibSp  Parch  \
0                            Braund, Mr. Owen Harris    1  22.0      1      0
1  Cumings, Mrs. John Bradley (Florence Briggs Th…    0  38.0      1      0
2                             Heikkinen, Miss. Laina    0  26.0      0      0
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)    0  35.0      1      0
4                           Allen, Mr. William Henry    1  35.0      0      0

             Ticket     Fare Cabin  Embarked
0         A/5 21171   7.2500   NaN         2
1          PC 17599  71.2833   C85         0
2  STON/O2. 3101282   7.9250   NaN         2
3            113803  53.1000  C123         2
4            373450   8.0500   NaN         2
```

[ ]:
```

```