# Lab 11 06-10-2022 k-NN telecus

October 6, 2022

## 1 Tele Customer Churn

[1]:
```
import numpy as np
import pandas as pd
!pip install scikit-learn
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: scikit-learn in
/opt/anaconda3/lib/python3.9/site-packages (1.0.2)
Requirement already satisfied: numpy>=1.14.6 in
/opt/anaconda3/lib/python3.9/site-packages (from scikit-learn) (1.21.5)
Requirement already satisfied: scipy>=1.1.0 in
/opt/anaconda3/lib/python3.9/site-packages (from scikit-learn) (1.7.3)
Requirement already satisfied: joblib>=0.11 in
/opt/anaconda3/lib/python3.9/site-packages (from scikit-learn) (1.1.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/opt/anaconda3/lib/python3.9/site-packages (from scikit-learn) (2.2.0)
```

[2]:
```
ds = pd.read_csv('Telco-Customer-Churn.csv')
print(ds)
```

```
      customerID  gender  SeniorCitizen Partner Dependents  tenure  \
0     7590-VHVEG  Female              0     Yes         No       1
1     5575-GNVDE    Male              0      No         No      34
2     3668-QPYBK    Male              0      No         No       2
3     7795-CFOCW    Male              0      No         No      45
4     9237-HQITU  Female              0      No         No       2
…          …     …             …      …        …     …
7038  6840-RESVB    Male              0     Yes        Yes      24
7039  2234-XADUH  Female              0     Yes        Yes      72
7040  4801-JZAZL  Female              0     Yes        Yes      11
7041  8361-LTMKD    Male              1     Yes         No       4
7042  3186-AJIEK    Male              0      No         No      66

     PhoneService     MultipleLines InternetService OnlineSecurity  …  \
0              No  No phone service             DSL             No  …
1             Yes                No             DSL            Yes  …
2             Yes                No             DSL            Yes  …
```

```
3              No  No phone service             DSL            Yes  …
4             Yes                 No     Fiber optic             No  …
…             …                   …              …             …   …
7038          Yes                Yes             DSL            Yes  …
7039          Yes                Yes     Fiber optic             No  …
7040           No  No phone service             DSL            Yes  …
7041          Yes                Yes     Fiber optic             No  …
7042          Yes                 No     Fiber optic            Yes  …

     DeviceProtection TechSupport StreamingTV StreamingMovies        Contract  \
0                  No          No          No              No  Month-to-month
1                 Yes          No          No              No        One year
2                  No          No          No              No  Month-to-month
3                 Yes         Yes          No              No        One year
4                  No          No          No              No  Month-to-month
…                  …           …           …               …              …
7038              Yes         Yes         Yes             Yes        One year
7039              Yes          No         Yes             Yes        One year
7040               No          No          No              No  Month-to-month
7041               No          No          No              No  Month-to-month
7042              Yes         Yes         Yes             Yes        Two year

     PaperlessBilling              PaymentMethod MonthlyCharges  TotalCharges  \
0                 Yes           Electronic check          29.85         29.85
1                  No               Mailed check          56.95        1889.5
2                 Yes               Mailed check          53.85        108.15
3                  No  Bank transfer (automatic)          42.30       1840.75
4                 Yes           Electronic check          70.70        151.65
…                 …                          …             …           …
7038              Yes               Mailed check          84.80        1990.5
7039              Yes    Credit card (automatic)         103.20        7362.9
7040              Yes           Electronic check          29.60        346.45
7041              Yes               Mailed check          74.40         306.6
7042              Yes  Bank transfer (automatic)         105.65        6844.5

     Churn
0       No
1       No
2      Yes
3       No
4      Yes
…      …
7038    No
7039    No
7040    No
7041   Yes
7042    No
```

```
[7043 rows x 21 columns]
```

[3]:
```
ds.drop('customerID', axis = 1,inplace=True)
```

[4]:
```
from sklearn.preprocessing import LabelEncoder
label_encoder_x= LabelEncoder()
```

[5]:
```
y = ds['Churn']
print(y)
```

```
0        No
1        No
2       Yes
3        No
4       Yes
       ...
7038     No
7039     No
7040     No
7041    Yes
7042     No
Name: Churn, Length: 7043, dtype: object
```

[6]:
```
ds.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 20 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   gender            7043 non-null   object
 1   SeniorCitizen     7043 non-null   int64
 2   Partner           7043 non-null   object
 3   Dependents        7043 non-null   object
 4   tenure            7043 non-null   int64
 5   PhoneService      7043 non-null   object
 6   MultipleLines     7043 non-null   object
 7   InternetService   7043 non-null   object
 8   OnlineSecurity    7043 non-null   object
 9   OnlineBackup      7043 non-null   object
 10  DeviceProtection  7043 non-null   object
 11  TechSupport       7043 non-null   object
 12  StreamingTV       7043 non-null   object
 13  StreamingMovies   7043 non-null   object
 14  Contract          7043 non-null   object
 15  PaperlessBilling  7043 non-null   object
 16  PaymentMethod     7043 non-null   object
 17  MonthlyCharges    7043 non-null   float64
```

```
 18  TotalCharges      7043 non-null   object
 19  Churn             7043 non-null   object
dtypes: float64(1), int64(2), object(17)
memory usage: 1.1+ MB
```

[7]:
```
ds['TotalCharges'] = pd.to_numeric(ds['TotalCharges'], errors='coerce')
ds.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 20 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   gender            7043 non-null   object
 1   SeniorCitizen     7043 non-null   int64
 2   Partner           7043 non-null   object
 3   Dependents        7043 non-null   object
 4   tenure            7043 non-null   int64
 5   PhoneService      7043 non-null   object
 6   MultipleLines     7043 non-null   object
 7   InternetService   7043 non-null   object
 8   OnlineSecurity    7043 non-null   object
 9   OnlineBackup      7043 non-null   object
 10  DeviceProtection  7043 non-null   object
 11  TechSupport       7043 non-null   object
 12  StreamingTV       7043 non-null   object
 13  StreamingMovies   7043 non-null   object
 14  Contract          7043 non-null   object
 15  PaperlessBilling  7043 non-null   object
 16  PaymentMethod     7043 non-null   object
 17  MonthlyCharges    7043 non-null   float64
 18  TotalCharges      7032 non-null   float64
 19  Churn             7043 non-null   object
dtypes: float64(2), int64(2), object(16)
memory usage: 1.1+ MB
```

[8]:
```
y = label_encoder_x.fit_transform(y)
print(y)
```

```
[0 0 1 … 0 1 0]
```

[9]:
```
x = ds.iloc[: , :-1].values
print(x)
```

```
[['Female' 0 'Yes' … 'Electronic check' 29.85 29.85]
 ['Male' 0 'No' … 'Mailed check' 56.95 1889.5]
 ['Male' 0 'No' … 'Mailed check' 53.85 108.15]
 …
```

```
['Female' 0 'Yes' … 'Electronic check' 29.6 346.45]
['Male' 1 'Yes' … 'Mailed check' 74.4 306.6]
['Male' 0 'No' … 'Bank transfer (automatic)' 105.65 6844.5]]
```

[15]:
```
ds.iloc[:, 0]= label_encoder_x.fit_transform(ds.iloc[:, 0])
ds.iloc[:, 1]= label_encoder_x.fit_transform(ds.iloc[:, 1])
ds.iloc[:, 2]= label_encoder_x.fit_transform(ds.iloc[:, 2])
ds.iloc[:, 3]= label_encoder_x.fit_transform(ds.iloc[:, 3])
ds.iloc[:, 4]= label_encoder_x.fit_transform(ds.iloc[:, 4])
ds.iloc[:, 5]= label_encoder_x.fit_transform(ds.iloc[:, 5])
ds.iloc[:, 6]= label_encoder_x.fit_transform(ds.iloc[:, 6])
ds.iloc[:, 7]= label_encoder_x.fit_transform(ds.iloc[:, 7])
ds.iloc[:, 8]= label_encoder_x.fit_transform(ds.iloc[:, 8])
ds.iloc[:, 9]= label_encoder_x.fit_transform(ds.iloc[:, 9])
ds.iloc[:, 10]= label_encoder_x.fit_transform(ds.iloc[:,10])
ds.iloc[:, 11]= label_encoder_x.fit_transform(ds.iloc[:, 11])
ds.iloc[:, 12]= label_encoder_x.fit_transform(ds.iloc[:, 12])
ds.iloc[:, 13]= label_encoder_x.fit_transform(ds.iloc[:, 13])
ds.iloc[:, 14]= label_encoder_x.fit_transform(ds.iloc[:, 14])
ds.iloc[:, 15]= label_encoder_x.fit_transform(ds.iloc[:, 15])
ds.iloc[:, 16]= label_encoder_x.fit_transform(ds.iloc[:, 16])
ds.iloc[:, 17]= label_encoder_x.fit_transform(ds.iloc[:, 17])
ds.iloc[:, 18]= label_encoder_x.fit_transform(ds.iloc[:, 18])
ds.iloc[:, 19]= label_encoder_x.fit_transform(ds.iloc[:, 19])
print(ds)
```

```
      gender  SeniorCitizen  Partner  Dependents  tenure  PhoneService  \
0          0              0        1           0       1             0
1          1              0        0           0      34             1
2          1              0        0           0       2             1
3          1              0        0           0      45             0
4          0              0        0           0       2             1
…        …              …        …           …      …             …
7038       1              0        1           1      24             1
7039       0              0        1           1      72             1
7040       0              0        1           1      11             0
7041       1              1        1           0       4             1
7042       1              0        0           0      66             1

      MultipleLines  InternetService  OnlineSecurity  OnlineBackup  \
0                 1                0               0             2
1                 0                0               2             0
2                 0                0               2             2
3                 1                0               2             0
4                 0                1               0             0
…               …                …               …             …
7038              2                0               2             0
7039              2                1               0             2
```

```
7040               1                 0               2               0
7041               2                 1               0               0
7042               0                 1               2               0

      DeviceProtection  TechSupport  StreamingTV  StreamingMovies  Contract  \
0                    0            0            0                0         0
1                    2            0            0                0         1
2                    0            0            0                0         0
3                    2            2            0                0         1
4                    0            0            0                0         0
…                  …            …            …                …         …
7038                 2            2            2                2         1
7039                 2            0            2                2         1
7040                 0            0            0                0         0
7041                 0            0            0                0         0
7042                 2            2            2                2         2

      PaperlessBilling  PaymentMethod  MonthlyCharges  TotalCharges  Churn
0                    1              2             142            74      0
1                    0              3             498          3624      0
2                    1              3             436           536      1
3                    0              0             266          3570      0
4                    1              2             729           674      1
…                  …              …               …             …     …
7038                 1              3             991          3700      0
7039                 1              1            1340          6304      0
7040                 1              2             137          1265      0
7041                 1              3             795          1157      1
7042                 1              0            1388          6150      0

[7043 rows x 20 columns]
```

[17]:
```
x = ds
one_hot_encoding_columns = ['MultipleLines', 'InternetService',␣
 ↪'OnlineSecurity', 'OnlineBackup', 'DeviceProtection',
                            'TechSupport', 'StreamingTV', 'StreamingMovies',␣
 ↪'Contract', 'PaymentMethod']

x = pd.get_dummies(x,columns = one_hot_encoding_columns)
print(x)
```

```
      gender  SeniorCitizen  Partner  Dependents  tenure  PhoneService  \
0          0              0        1           0       1             0
1          1              0        0           0      34             1
2          1              0        0           0       2             1
3          1              0        0           0      45             0
4          0              0        0           0       2             1
…        …              …        …           …       …             …
```

```
7038       1              0        1           1      24             1
7039       0              0        1           1      72             1
7040       0              0        1           1      11             0
7041       1              1        1           0       4             1
7042       1              0        0           0      66             1

      PaperlessBilling  MonthlyCharges  TotalCharges  Churn  …  \
0                    1             142            74      0  …
1                    0             498          3624      0  …
2                    1             436           536      1  …
3                    0             266          3570      0  …
4                    1             729           674      1  …
…                  …               …             …     …  …
7038                 1             991          3700      0  …
7039                 1            1340          6304      0  …
7040                 1             137          1265      0  …
7041                 1             795          1157      1  …
7042                 1            1388          6150      0  …

      StreamingMovies_0  StreamingMovies_1  StreamingMovies_2  Contract_0  \
0                     1                  0                  0           1
1                     1                  0                  0           0
2                     1                  0                  0           1
3                     1                  0                  0           0
4                     1                  0                  0           1
…                   …                  …                  …         …
7038                  0                  0                  1           0
7039                  0                  0                  1           0
7040                  1                  0                  0           1
7041                  1                  0                  0           1
7042                  0                  0                  1           0

      Contract_1  Contract_2  PaymentMethod_0  PaymentMethod_1  \
0              0           0                0                0
1              1           0                0                0
2              0           0                0                0
3              1           0                1                0
4              0           0                0                0
…            …           …                …              …
7038           1           0                0                0
7039           1           0                0                1
7040           0           0                0                0
7041           0           0                0                0
7042           0           1                1                0

      PaymentMethod_2  PaymentMethod_3
0                   1                0
1                   0                1
```

```
2                    0                    1
3                    0                    0
4                    1                    0
…                    …                    …
7038                 0                    1
7039                 0                    0
7040                 1                    0
7041                 0                    1
7042                 0                    0

[7043 rows x 41 columns]
```

[18]:
```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test= train_test_split(x, y, test_size= 0.25,␣
 ↪random_state=0)
```

[19]:
```
from sklearn.preprocessing import MinMaxScaler
st_x= MinMaxScaler()
x_train= st_x.fit_transform(x_train)
x_test=st_x.fit_transform(x_test)
print(x_train)
```

```
[[0. 0. 0. … 1. 0. 0.]
 [1. 1. 0. … 0. 0. 0.]
 [1. 0. 0. … 0. 0. 1.]
 …
 [1. 0. 1. … 0. 0. 1.]
 [1. 1. 0. … 0. 1. 0.]
 [0. 0. 0. … 1. 0. 0.]]
```

[20]:
```
from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n_neighbors = 5)
classifier.fit(x_train,y_train)
```

[20]:
```
KNeighborsClassifier()
```

[22]:
```
y_pred= classifier.predict(x_test)
print(y_pred)
```

```
[0 0 0 … 0 0 0]
```

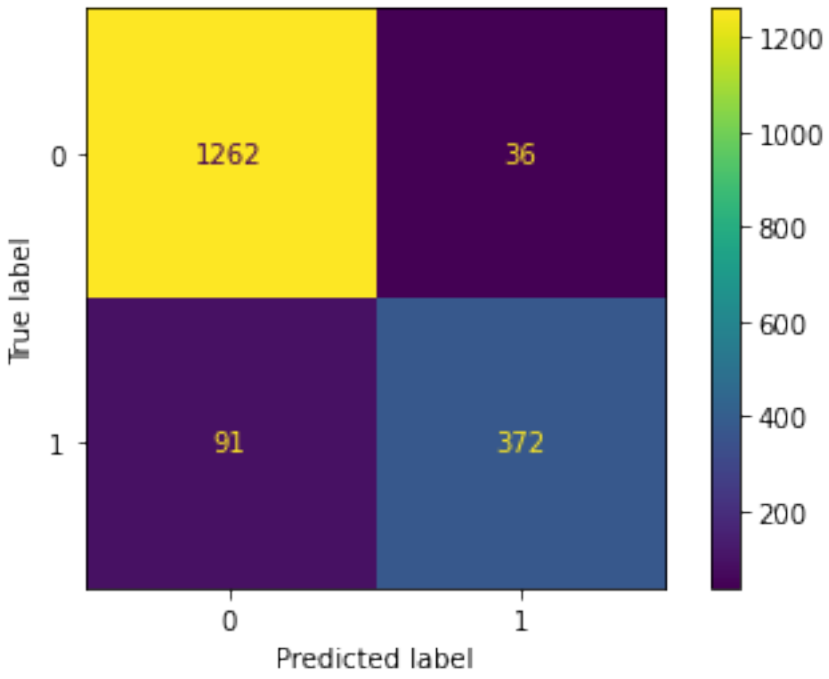
[23]:
```
from sklearn.metrics import confusion_matrix
cm= confusion_matrix (y_test, y_pred,labels=classifier.classes_)
print(cm)
```

```
[[1262   36]
 [  91  372]]
```

[24]:
```
from sklearn.metrics import ConfusionMatrixDisplay
disp = ConfusionMatrixDisplay(confusion_matrix=cm,
display_labels=classifier.classes_)

disp.plot()
```

[24]:
```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at
0x7f2ce28f9f10>
```




[25]:
```
training_score = classifier.score(x_train, y_train)
test_score = classifier.score(x_test, y_test)
print(training_score)
print(test_score)
```

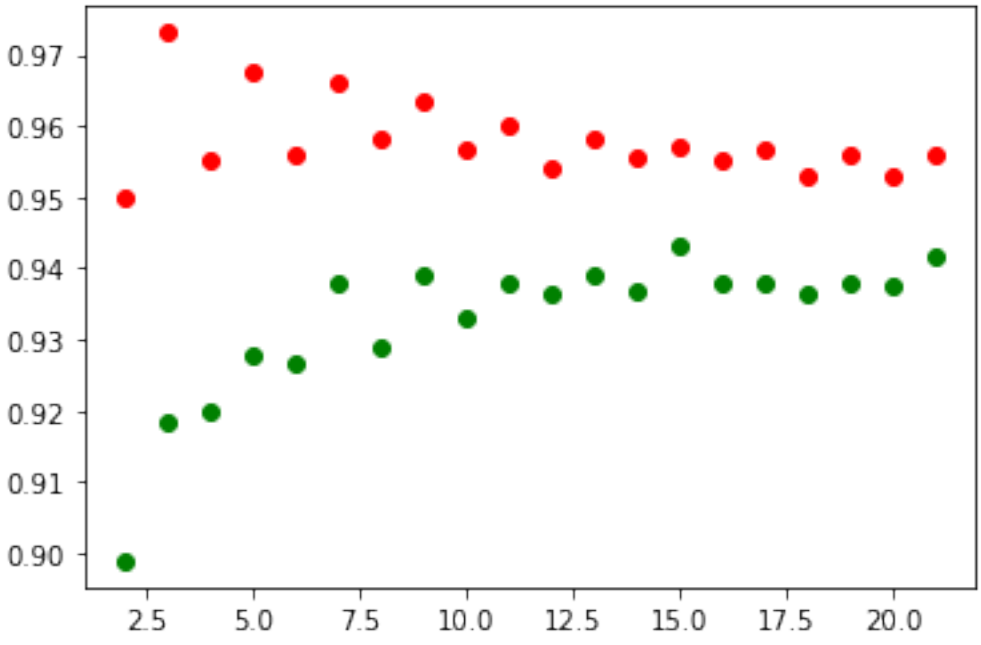
```
0.9674365770541462
0.9278818852924475
```

[26]:
```
K = []
training = []
test = []
scores = {}
for k in range(2, 22):
    clf = KNeighborsClassifier(n_neighbors = k)
    clf.fit(x_train, y_train)
    training_score = clf.score(x_train, y_train)
```

```
    test_score = clf.score(x_test, y_test)
    K.append(k)
    training.append(training_score)
    test.append(test_score)
    scores[k] = [training_score, test_score]
for keys, values in scores.items():
    print(keys, ':', values)
```

```
2 : [0.9500189322226429, 0.8989210675752414]
3 : [0.9731162438470277, 0.9182282793867121]
4 : [0.9551306323362363, 0.919931856899489]
5 : [0.9674365770541462, 0.9278818852924475]
6 : [0.9560772434683832, 0.9267461669505963]
7 : [0.9661113214691405, 0.9381033503691084]
8 : [0.9583491101855358, 0.9290176036342986]
9 : [0.9634608102991291, 0.9392390687109596]
10 : [0.9566452101476713, 0.9329926178307779]
11 : [0.9600530102234002, 0.9381033503691084]
12 : [0.95399469897766, 0.9363997728563316]
13 : [0.9583491101855358, 0.9392390687109596]
14 : [0.955509276789095, 0.9369676320272572]
15 : [0.9570238546005301, 0.9432140829074389]
16 : [0.9553199545626656, 0.9381033503691084]
17 : [0.9568345323741008, 0.9381033503691084]
18 : [0.9528587656190837, 0.9363997728563316]
19 : [0.9558879212419538, 0.9381033503691084]
20 : [0.9530480878455131, 0.9375354911981828]
21 : [0.9558879212419538, 0.9415105053946621]
```

[27]:
```
import matplotlib.pyplot as plt
plt.scatter(K, training, color ='r')
plt.scatter(K, test, color ='g')
plt.show()
```


0.97
0.96
0.95
0.94
0.93
0.92
0.91
0.90
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0