Рубежный контроль №2

Якубов Артём

Группа ИУ5-63Б

Вариант 24

Задача. Для заданного набора данных (по Вашему варианту) постройте модели классификации или регрессии (в зависимости от конкретной задачи, рассматриваемой в наборе данных). Для построения моделей используйте методы 1 и 2 (по варианту для Вашей группы). Оцените качество моделей на основе подходящих метрик качества (не менее двух метрик). Какие метрики качества Вы использовали и почему? Какие выводы Вы можете сделать о качестве построенных моделей? Для построения моделей необходимо выполнить требуемую предобработку данных: заполнение пропусков, кодирование категориальных признаков, и т.д.

Методы для ИУ5-63Б. Метод №1: "Дерево решений". Метод №2: "Случайный лес".

Импорт библиотек и первичное исследование датасета

In [1]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.impute import SimpleImputer
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, balanced_accuracy_score
from sklearn.metrics import precision_score, recall_score, f1_score, classification_re
from sklearn.metrics import mean_absolute_error, mean_squared_error, mean_squared_log_
from sklearn.preprocessing import LabelEncoder, OneHotEncoder
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.preprocessing import MinMaxScaler, StandardScaler, Normalizer
from sklearn.metrics import roc_auc_score
%matplotlib inline
sns.set(style="ticks")
```

In [2]:

```
data = pd.read_csv('telecom_users.csv', sep=",")
```

In [3]:

```
data.shape
```

Out[3]:

```
(5986, 22)
```

In [4]:

```
data.info
```

Out[4]:

```
<bound method DataFrame.info of       Unnamed: 0  customerID  gender  SeniorCitizen Pa
rtner Dependents  \
0           1869  7010-BRBUU    Male              0     Yes        Yes
1           4528  9688-YGXVR  Female              0      No         No
2           6344  9286-DOJGF  Female              1     Yes         No
3           6739  6994-KERXL    Male              0      No         No
4            432  2181-UAESM    Male              0      No         No
...          ...         ...     ...            ...     ...        ...
5981        3772  0684-AOSIH    Male              0     Yes         No
5982        5191  5982-PSMKW  Female              0     Yes        Yes
5983        5226  8044-BGWPI    Male              0     Yes        Yes
5984        5390  7450-NWRTR    Male              1      No         No
5985         860  4795-UXVCJ    Male              0      No         No

      tenure PhoneService MultipleLines InternetService  ...  \
0         72          Yes           Yes              No  ...
1         44          Yes            No     Fiber optic  ...
2         38          Yes           Yes     Fiber optic  ...
3          4          Yes            No             DSL  ...
4          2          Yes            No             DSL  ...
...      ...          ...           ...             ...  ...
5981       1          Yes            No     Fiber optic  ...
5982      23          Yes           Yes             DSL  ...
5983      12          Yes            No              No  ...
5984      12          Yes           Yes     Fiber optic  ...
5985      26          Yes            No              No  ...

         DeviceProtection          TechSupport          StreamingTV  \
0     No internet service  No internet service  No internet service
1                     Yes                   No                  Yes
2                      No                   No                   No
3                      No                   No                   No
4                     Yes                   No                   No
...                   ...                  ...                  ...
5981                   No                   No                  Yes
5982                  Yes                  Yes                  Yes
5983  No internet service  No internet service  No internet service
5984                  Yes                   No                  Yes
5985  No internet service  No internet service  No internet service

          StreamingMovies        Contract PaperlessBilling  \
0     No internet service        Two year               No
1                      No  Month-to-month              Yes
2                      No  Month-to-month              Yes
3                     Yes  Month-to-month              Yes
4                      No  Month-to-month               No
...                   ...             ...              ...
5981                  Yes  Month-to-month              Yes
5982                  Yes        Two year              Yes
5983  No internet service  Month-to-month              Yes
5984                  Yes  Month-to-month              Yes
5985  No internet service        One year               No

                  PaymentMethod MonthlyCharges TotalCharges  Churn
0       Credit card (automatic)          24.10      1734.65     No
1       Credit card (automatic)          88.15       3973.2     No
2     Bank transfer (automatic)          74.95      2869.85    Yes
3              Electronic check          55.90        238.5     No
4              Electronic check          53.45        119.5     No
...                         ...            ...          ...    ...
5981           Electronic check          95.00           95    Yes
5982    Credit card (automatic)          91.10       2198.3     No
5983           Electronic check          21.15       306.05     No
5984           Electronic check          99.45      1200.15    Yes
5985    Credit card (automatic)          19.80        457.3     No

[5986 rows x 22 columns]>
```

In [5]:

```
data.isnull().sum()
```

Out[5]:

```
Unnamed: 0          0
customerID          0
gender              0
SeniorCitizen       0
Partner             0
Dependents          0
tenure              0
PhoneService        0
MultipleLines       0
InternetService     0
OnlineSecurity      0
OnlineBackup        0
DeviceProtection    0
TechSupport         0
StreamingTV         0
StreamingMovies     0
Contract            0
PaperlessBilling    0
PaymentMethod       0
MonthlyCharges      0
TotalCharges        0
Churn               0
dtype: int64
```

In [6]:

```
data.dtypes
```

Out[6]:

```
Unnamed: 0            int64
customerID           object
gender               object
SeniorCitizen         int64
Partner              object
Dependents           object
tenure                int64
PhoneService         object
MultipleLines        object
InternetService      object
OnlineSecurity       object
OnlineBackup         object
DeviceProtection     object
TechSupport          object
StreamingTV          object
StreamingMovies      object
Contract             object
PaperlessBilling     object
PaymentMethod        object
MonthlyCharges      float64
TotalCharges         object
Churn                object
dtype: object
```

Кодирование категориальных признаков числовыми

In [7]:

```
cat_cols = ['gender', 'Partner', 'Dependents', 'PhoneService', 'MultipleLines', 'Inte
```

In [8]:

```
for col in data.columns:
    dt = str(data[col].dtype)
    if dt == 'object':
        temp_un = data[col].nunique()
        print('Колонка {}. Тип данных {}. Количество уникальных значений {}.'.format(c
```

```
Колонка customerID. Тип данных object. Количество уникальных значений 5986.
Колонка gender. Тип данных object. Количество уникальных значений 2.
Колонка Partner. Тип данных object. Количество уникальных значений 2.
Колонка Dependents. Тип данных object. Количество уникальных значений 2.
Колонка PhoneService. Тип данных object. Количество уникальных значений 2.
Колонка MultipleLines. Тип данных object. Количество уникальных значений 3.
Колонка InternetService. Тип данных object. Количество уникальных значений 3.
Колонка OnlineSecurity. Тип данных object. Количество уникальных значений 3.
Колонка OnlineBackup. Тип данных object. Количество уникальных значений 3.
Колонка DeviceProtection. Тип данных object. Количество уникальных значений 3.
Колонка TechSupport. Тип данных object. Количество уникальных значений 3.
Колонка StreamingTV. Тип данных object. Количество уникальных значений 3.
Колонка StreamingMovies. Тип данных object. Количество уникальных значений 3.
Колонка Contract. Тип данных object. Количество уникальных значений 3.
Колонка PaperlessBilling. Тип данных object. Количество уникальных значений 2.
Колонка PaymentMethod. Тип данных object. Количество уникальных значений 4.
Колонка TotalCharges. Тип данных object. Количество уникальных значений 5611.
Колонка Churn. Тип данных object. Количество уникальных значений 2.
```

In [9]:

```
le = LabelEncoder()
data.loc[:, 'customerID'] = le.fit_transform(data['customerID'])
data['customerID'].head()
```

Out[9]:

```
0    4252
1    5806
2    5577
3    4243
4    1270
Name: customerID, dtype: int32
```

In [10]:

```
data.loc[:, 'TotalCharges'] = le.fit_transform(data['TotalCharges'])
data['TotalCharges'].head()
```

Out[10]:

```
0    1066
1    2901
2    2126
3    1733
4     308
Name: TotalCharges, dtype: int32
```

In [11]:

```
one_hot = pd.get_dummies(data[cat_cols].astype(str))
one_hot.head()
```

Out[11]:

	gender_Female	gender_Male	Partner_No	Partner_Yes	Dependents_No	Dependents_Yes	PhoneService_No
0	0	1	0	1	0	1	0
1	1	0	1	0	1	0	0
2	1	0	0	1	1	0	0
3	0	1	1	0	1	0	0
4	0	1	1	0	1	0	0

5 rows × 43 columns

In [12]:

```
data = data.join(one_hot)
data.drop(columns=cat_cols, inplace=True)
```

In [13]:

```
data.shape
```

Out[13]:

```
(5986, 49)
```

In [14]:

```
data.head()
```

Out[14]:

	Unnamed: 0	customerID	SeniorCitizen	tenure	MonthlyCharges	TotalCharges	gender_Female	gender_Male
0	1869	4252	0	72	24.10	1066	0	1
1	4528	5806	0	44	88.15	2901	1	0
2	6344	5577	1	38	74.95	2126	1	0
3	6739	4243	0	4	55.90	1733	0	1
4	432	1270	0	2	53.45	308	0	1

5 rows × 49 columns

Разделение и обучение выборки

In [15]:

```
parts = np.split(data, [48,49], axis=1)
X = parts[0]
Y = parts[1]
print('Входные данные:\n\n', X.head(), '\n\nВыходные данные:\n\n', Y.head())
```

```
Входные данные:

    Unnamed: 0  customerID  SeniorCitizen  tenure  MonthlyCharges  \
0      1869.0      4252.0            0.0    72.0           24.10
1      4528.0      5806.0            0.0    44.0           88.15
2      6344.0      5577.0            1.0    38.0           74.95
3      6739.0      4243.0            0.0     4.0           55.90
4       432.0      1270.0            0.0     2.0           53.45

   TotalCharges  gender_Female  gender_Male  Partner_No  Partner_Yes  ...  \
0        1066.0            0.0          1.0         0.0          1.0  ...
1        2901.0            1.0          0.0         1.0          0.0  ...
2        2126.0            1.0          0.0         0.0          1.0  ...
3        1733.0            0.0          1.0         1.0          0.0  ...
4         308.0            0.0          1.0         1.0          0.0  ...

   Contract_Month-to-month  Contract_One year  Contract_Two year  \
0                      0.0                0.0                1.0
1                      1.0                0.0                0.0
2                      1.0                0.0                0.0
3                      1.0                0.0                0.0
4                      1.0                0.0                0.0

   PaperlessBilling_No  PaperlessBilling_Yes  \
0                  1.0                   0.0
1                  0.0                   1.0
2                  0.0                   1.0
3                  0.0                   1.0
4                  1.0                   0.0

   PaymentMethod_Bank transfer (automatic)  \
0                                      0.0
1                                      0.0
2                                      1.0
3                                      0.0
4                                      0.0

   PaymentMethod_Credit card (automatic)  PaymentMethod_Electronic check  \
0                                    1.0                             0.0
1                                    1.0                             0.0
2                                    0.0                             0.0
3                                    0.0                             1.0
4                                    0.0                             1.0

   PaymentMethod_Mailed check  Churn_No
0                         0.0       1.0
1                         0.0       1.0
2                         0.0       0.0
3                         0.0       1.0
4                         0.0       1.0

[5 rows x 48 columns]

Выходные данные:

    Churn_Yes
0        0.0
1        0.0
2        1.0
3        0.0
4        0.0
```

In [16]:

```
X_train,  X_test,  Y_train,  Y_test = train_test_split(X,  Y, test_size=0.5)
```

In [17]:

```
print('Входные параметры обучающей выборки:\n\n',X_train.head(),
      '\n\nВходные параметры тестовой выборки:\n\n', X_test.head(),
      '\n\nВыходные параметры обучающей выборки:\n\n', Y_train.head(),
      '\n\nВыходные параметры тестовой выборки:\n\n', Y_test.head())
```

```
Входные параметры обучающей выборки:

       Unnamed: 0  customerID  SeniorCitizen  tenure  MonthlyCharges  \
1812      4240.0       923.0            0.0     8.0           94.00
1288      5312.0      5735.0            0.0     1.0           51.25
5859      4514.0      2369.0            0.0    28.0           80.60
825       3105.0      2625.0            0.0    26.0           49.15
2196      6117.0      1973.0            0.0    69.0           19.30

      TotalCharges  gender_Female  gender_Male  Partner_No  Partner_Yes  ...  \
1812        5060.0            0.0          1.0         1.0          0.0  ...
1288        3696.0            0.0          1.0         1.0          0.0  ...
5859        1613.0            1.0          0.0         1.0          0.0  ...
825          376.0            0.0          1.0         1.0          0.0  ...
2196         703.0            0.0          1.0         1.0          0.0  ...

      Contract_Month-to-month  Contract_One year  Contract_Two year  \
1812                      1.0                0.0                0.0
1288                      1.0                0.0                0.0
5859                      1.0                0.0                0.0
825                       1.0                0.0                0.0
2196                      0.0                0.0                1.0

      PaperlessBilling_No  PaperlessBilling_Yes  \
1812                  0.0                   1.0
1288                  1.0                   0.0
5859                  0.0                   1.0
825                   1.0                   0.0
2196                  0.0                   1.0

      PaymentMethod_Bank transfer (automatic)  \
1812                                      0.0
1288                                      0.0
5859                                      0.0
825                                       1.0
2196                                      1.0

      PaymentMethod_Credit card (automatic)  PaymentMethod_Electronic check  \
1812                                    0.0                             1.0
1288                                    0.0                             0.0
5859                                    0.0                             1.0
825                                     0.0                             0.0
2196                                    0.0                             0.0

      PaymentMethod_Mailed check  Churn_No
1812                         0.0       0.0
1288                         1.0       0.0
5859                         0.0       1.0
825                          0.0       1.0
2196                         0.0       1.0

[5 rows x 48 columns]

Входные параметры тестовой выборки:

       Unnamed: 0  customerID  SeniorCitizen  tenure  MonthlyCharges  \
3893      5220.0      1355.0            0.0    29.0           58.00
3810       916.0      2368.0            1.0    23.0           75.40
731        605.0      2830.0            0.0    72.0           86.65
5075      1331.0      4545.0            0.0     1.0           25.15
5265      6381.0       541.0            0.0     1.0           19.55

      TotalCharges  gender_Female  gender_Male  Partner_No  Partner_Yes  ...  \
3893        1065.0            1.0          0.0         1.0          0.0  ...
3810         958.0            1.0          0.0         1.0          0.0  ...
731         4273.0            0.0          1.0         0.0          1.0  ...
5075        1844.0            1.0          0.0         1.0          0.0  ...
5265        1272.0            0.0          1.0         1.0          0.0  ...

      Contract_Month-to-month  Contract_One year  Contract_Two year  \
3893                      0.0                1.0                0.0
3810                      1.0                0.0                0.0
731                       0.0                0.0                1.0
5075                      1.0                0.0                0.0
5265                      1.0                0.0                0.0

      PaperlessBilling_No  PaperlessBilling_Yes  \
3893                  1.0                   0.0
3810                  0.0                   1.0
731                   1.0                   0.0
5075                  0.0                   1.0
5265                  1.0                   0.0

      PaymentMethod_Bank transfer (automatic)  \
3893                                      0.0
3810                                      1.0
731                                       0.0
5075                                      0.0
5265                                      0.0

      PaymentMethod_Credit card (automatic)  PaymentMethod_Electronic check  \
3893                                    1.0                             0.0
3810                                    0.0                             0.0
731                                     1.0                             0.0
5075                                    0.0                             0.0
5265                                    0.0                             0.0

      PaymentMethod_Mailed check  Churn_No
3893                         0.0       1.0
3810                         0.0       1.0
731                          0.0       1.0
5075                         1.0       1.0
5265                         1.0       1.0

[5 rows x 48 columns]

Выходные параметры обучающей выборки:

       Churn_Yes
1812        1.0
1288        1.0
5859        0.0
825         0.0
2196        0.0

Выходные параметры тестовой выборки:

       Churn_Yes
3893        0.0
3810        0.0
731         0.0
5075        0.0
5265        0.0
```

In [18]:

```
dtc = DecisionTreeRegressor(random_state=1).fit(X_train, Y_train.values.ravel())
data_test_predicted_dtc = dtc.predict(X_test)
```

In [19]:

```
random_forest = RandomForestRegressor(random_state=1).fit(X_train, Y_train.values.rave
data_test_predicted_rf = random_forest.predict(X_test)
```

In [25]:

```
print('Метрика MSE:\nДерево решений: {}\nСлучайный лес: {}'.format(mean_squared_error
```

```
Метрика MSE:
Дерево решений: 0.0
Случайный лес: 0.0
```

In [22]:

```
print('Метрика R\u00B2:\nДерево решений: {}\nСлучайный лес: {}'.format(r2_score(Y_test
```

```
Метрика R²:
Дерево решений: 1.0
Случайный лес: 1.0
```