Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ Информатика и системы управления

КАФЕДРА Системы обработки информации и управления

Отчёт по рубежному контролю №2

По дисциплине:
«Технологии машинного обучения»

Выполнил:

Студент группы ИУ5-63Б **Ваксина И.Р.**

(Подпись, дата) (Фамилия И.О.)

Проверил:

Гапанюк Ю.Е.

(Подпись, дата) (Фамилия И.О.)

Москва, 2021

Задание.

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.

Набор данных:

1. hhttps://www.kaggle.com/brsdincer/star-type-classification

Импорт библиотек

In [1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from pandas.plotting import scatter_matrix
from sklearn.model_selection import train_test_split
import warnings
warnings.filterwarnings('ignore')
sns.set(style="ticks")
%matplotlib inline
```

In [2]:
```
data = pd.read_csv('Stars.csv')
```

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

Out[3]:

	Temperature	L	R	A_M	Color	Spectral_Class	Type
0	3068	0.002400	0.1700	16.12	Red	M	0
1	3042	0.000500	0.1542	16.60	Red	M	0
2	2600	0.000300	0.1020	18.70	Red	M	0
3	2800	0.000200	0.1600	16.65	Red	M	0
4	1939	0.000138	0.1030	20.06	Red	M	0

In [4]:
```
data.dtypes
```

Out[4]:
```
Temperature         int64
L                 float64
R                 float64
A_M               float64
Color              object
Spectral_Class     object
Type                int64
dtype: object
```

In [5]:
```
data.drop(['Color','Spectral_Class','Color'], axis = 1, inplace = True)
```

In [6]:
```
data.isnull().sum()
# проверим есть ли пропущенные значения
```

Out[6]:
```
Temperature    0
L              0
R              0
A_M            0
Type           0
dtype: int64
```

In [7]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 240 entries, 0 to 239
Data columns (total 5 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   Temperature  240 non-null    int64  
 1   L            240 non-null    float64
 2   R            240 non-null    float64
 3   A_M          240 non-null    float64
 4   Type         240 non-null    int64  
dtypes: float64(3), int64(2)
memory usage: 9.5 KB
```

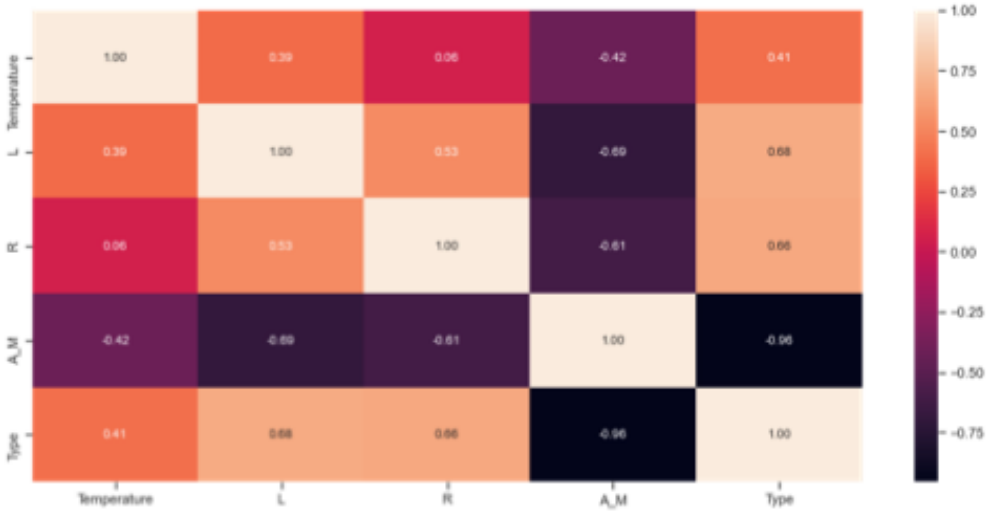
In [8]:
```
data.head()
```

Out[8]:

	Temperature	L	R	A_M	Type
0	3068	0.002400	0.1700	16.12	0
1	3042	0.000500	0.1542	16.60	0
2	2600	0.000300	0.1020	18.70	0
3	2800	0.000200	0.1600	16.65	0
4	1939	0.000138	0.1030	20.06	0

In [9]:
```
#Построим корреляционную матрицу
fig, ax = plt.subplots(figsize=(15,7))
sns.heatmap(data.corr(method='pearson'), ax=ax, annot=True, fmt='.2f')
```

Out[9]:
```
<AxesSubplot:>
```



In [10]:
```
data['L'] = data['L'].astype(int)
X = data.drop(['L'], axis = 1)
Y = data.L
print('Входные данные:\n\n', X.head(), '\n\nВыходные данные:\n\n', Y.head(
```

Входные данные:

	Temperature	R	A_M	Type
0	3068	0.1700	16.12	0
1	3042	0.1542	16.60	0
2	2600	0.1020	18.70	0
3	2800	0.1600	16.65	0
4	1939	0.1030	20.06	0

Выходные данные:

```
0    0
1    0
2    0
3    0
4    0
Name: L, dtype: int64
```

In [11]:
```
X_train,  X_test,  Y_train,  Y_test = train_test_split(X,  Y, random_state
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:

	Temperature	R	A_M	Type
5	2840	0.110	16.980	0
22	7220	0.011	14.230	2
199	3463	0.675	14.776	1
97	7720	1.340	2.440	3
12	3134	0.196	13.210	1

Входные параметры тестовой выборки:

	Temperature	R	A_M	Type
109	33421	67.000	-5.79	4
71	3607	0.380	10.12	1
37	6380	0.980	2.93	3
74	3550	0.291	10.89	1
108	24345	57.000	-6.24	4

Выходные параметры обучающей выборки:

```
5      0
22     0
199    0
97     7
12     0
Name: L, dtype: int64
```

Выходные параметры тестовой выборки:

```
109    352000
71          0
37          1
74          0
108    142000
Name: L, dtype: int64
```

In [12]:
```
from sklearn.tree import DecisionTreeClassifier, DecisionTreeRegressor, exp
from sklearn.tree import export_graphviz
from sklearn import tree
import re
```

In [13]:
```
clf = tree.DecisionTreeClassifier()
clf = clf.fit(X, Y)
from IPython.core.display import HTML
from sklearn.tree.export import export_text
tree_rules = export_text(clf, feature_names=list(X.columns))
HTML('<pre>' + tree_rules + '</pre>')
```

Out[13]:
```
|--- A_M <= 4.57
|   |--- R <= 1.23
|   |   |--- class: 1
|   |--- R >  1.23
|   |   |--- Temperature <= 3830.00
|   |   |   |--- Temperature <= 3766.00
|   |   |   |   |--- Temperature <= 3544.00
|   |   |   |   |   |--- A_M <= -6.71
|   |   |   |   |   |   |--- A_M <= -11.55
|   |   |   |   |   |   |   |--- class: 263000
|   |   |   |   |   |   |--- A_M >  -11.55
|   |   |   |   |   |   |   |--- Type <= 4.50
|   |   |   |   |   |   |   |   |--- class: 195000
|   |   |   |   |   |   |   |--- Type >  4.50
|   |   |   |   |   |   |   |   |--- R <= 1286.50
|   |   |   |   |   |   |   |   |   |--- class: 174000
|   |   |   |   |   |   |   |   |--- R >  1286.50
|   |   |   |   |   |   |   |   |   |--- R <= 1354.50
|   |   |   |   |   |   |   |   |   |   |--- class: 100000
|   |   |   |   |   |   |   |   |   |--- R >  1354.50
|   |   |   |   |   |   |   |   |   |   |--- R <= 1453.00
|   |   |   |   |   |   |   |   |   |   |   |--- class: 138000
|   |   |   |   |   |   |   |   |   |   |--- R >  1453.00
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch o
f depth 3
|   |   |   |   |   |--- A_M >  -6.71
|   |   |   |   |   |   |--- A_M <= -6.11
|   |   |   |   |   |   |   |--- class: 340000
|   |   |   |   |   |   |--- A_M >  -6.11
|   |   |   |   |   |   |   |--- A_M <= -6.01
|   |   |   |   |   |   |   |   |--- class: 150000
|   |   |   |   |   |   |   |--- A_M >  -6.01
|   |   |   |   |   |   |   |   |--- class: 280000
|   |   |   |   |--- Temperature >  3544.00
|   |   |   |   |   |--- Temperature <= 3620.00
|   |   |   |   |   |   |--- A_M <= -7.64
|   |   |   |   |   |   |   |--- R <= 1257.00
|   |   |   |   |   |   |   |   |--- Temperature <= 3602.50
|   |   |   |   |   |   |   |   |   |--- class: 240000
|   |   |   |   |   |   |   |   |--- Temperature >  3602.50
|   |   |   |   |   |   |   |   |   |--- class: 126000
|   |   |   |   |   |   |   |--- R >  1257.00
|   |   |   |   |   |   |   |   |--- R <= 1594.00
|   |   |   |   |   |   |   |   |   |--- A_M <= -10.94
|   |   |   |   |   |   |   |   |   |   |--- class: 145000
|   |   |   |   |   |   |   |   |   |--- A_M >  -10.94
|   |   |   |   |   |   |   |   |   |   |--- A_M <= -9.28
```

```
|   |   |   |   |   |   |   |   |   |   |   |--- class: 132000
|   |   |   |   |   |   |   |   |   |   |--- A_M >  -9.28
|   |   |   |   |   |   |   |   |   |   |   |--- class: 145000
|   |   |   |   |   |   |   |   |   |--- R >  1594.00
|   |   |   |   |   |   |   |   |   |--- class: 200000
|   |   |   |   |   |   |   |--- A_M >  -7.64
|   |   |   |   |   |   |   |   |--- A_M <= -5.92
|   |   |   |   |   |   |   |   |   |--- Type <= 4.50
|   |   |   |   |   |   |   |   |   |   |--- Temperature <= 3587.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: 123000
|   |   |   |   |   |   |   |   |   |   |--- Temperature >  3587.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: 320000
|   |   |   |   |   |   |   |   |   |--- Type >  4.50
|   |   |   |   |   |   |   |   |   |   |--- class: 320000
|   |   |   |   |   |   |   |   |--- A_M >  -5.92
|   |   |   |   |   |   |   |   |   |--- class: 200000
|   |   |   |   |   |   |--- Temperature >  3620.00
|   |   |   |   |   |   |   |--- R <= 480.00
|   |   |   |   |   |   |   |   |--- class: 184000
|   |   |   |   |   |   |   |--- R >  480.00
|   |   |   |   |   |   |   |   |--- A_M <= -11.58
|   |   |   |   |   |   |   |   |   |--- class: 363000
|   |   |   |   |   |   |   |   |--- A_M >  -11.58
|   |   |   |   |   |   |   |   |   |--- A_M <= -10.74
|   |   |   |   |   |   |   |   |   |   |--- class: 209000
|   |   |   |   |   |   |   |   |   |--- A_M >  -10.74
|   |   |   |   |   |   |   |   |   |   |--- R <= 1068.00
|   |   |   |   |   |   |   |   |   |   |   |--- class: 74000
|   |   |   |   |   |   |   |   |   |   |--- R >  1068.00
|   |   |   |   |   |   |   |   |   |   |   |--- Temperature <= 3699.50
|   |   |   |   |   |   |   |   |   |   |   |   |--- class: 310000
|   |   |   |   |   |   |   |   |   |   |   |--- Temperature >  3699.50
|   |   |   |   |   |   |   |   |   |   |   |   |--- truncated branch o
f depth 2
|   |   |   |--- Temperature >  3766.00
|   |   |   |   |--- class: 200000
|   |   |--- Temperature >  3830.00
|   |   |   |--- Temperature <= 26256.50
|   |   |   |   |--- Type <= 3.50
|   |   |   |   |   |--- A_M <= -3.89
|   |   |   |   |   |   |--- class: 14500
|   |   |   |   |   |--- A_M >  -3.89
|   |   |   |   |   |   |--- A_M <= -3.74
|   |   |   |   |   |   |   |--- class: 14520
|   |   |   |   |   |   |--- A_M >  -3.74
|   |   |   |   |   |   |   |--- A_M <= -3.54
|   |   |   |   |   |   |   |   |--- class: 12450
|   |   |   |   |   |   |   |--- A_M >  -3.54
|   |   |   |   |   |   |   |   |--- Temperature <= 7886.00
|   |   |   |   |   |   |   |   |   |--- class: 7
|   |   |   |   |   |   |   |   |--- Temperature >  7886.00
|   |   |   |   |   |   |   |   |   |--- A_M <= -3.36
|   |   |   |   |   |   |   |   |   |   |--- class: 4720
|   |   |   |   |   |   |   |   |   |--- A_M >  -3.36
```

```
|   |   |   |   |   |   |   |   |   |   |--- R <= 1.85
|   |   |   |   |   |   |   |   |   |   |   |--- class: 8
|   |   |   |   |   |   |   |   |   |   |--- R >  1.85
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch o
f depth 16
|   |   |   |   |--- Type >  3.50
|   |   |   |   |   |--- R <= 18.00
|   |   |   |   |   |   |--- class: 112000
|   |   |   |   |   |--- R >  18.00
|   |   |   |   |   |   |--- Temperature <= 3924.50
|   |   |   |   |   |   |   |--- class: 272000
|   |   |   |   |   |   |--- Temperature >  3924.50
|   |   |   |   |   |   |   |--- Temperature <= 4151.00
|   |   |   |   |   |   |   |   |--- class: 282000
|   |   |   |   |   |   |   |--- Temperature >  4151.00
|   |   |   |   |   |   |   |   |--- Type <= 4.50
|   |   |   |   |   |   |   |   |   |--- A_M <= -7.40
|   |   |   |   |   |   |   |   |   |   |--- class: 224780
|   |   |   |   |   |   |   |   |   |--- A_M >  -7.40
|   |   |   |   |   |   |   |   |   |   |--- R <= 24.50
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch o
f depth 2
|   |   |   |   |   |   |   |   |   |   |--- R >  24.50
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch o
f depth 19
|   |   |   |   |   |   |   |   |--- Type >  4.50
|   |   |   |   |   |   |   |   |   |--- Temperature <= 5568.50
|   |   |   |   |   |   |   |   |   |   |--- class: 630000
|   |   |   |   |   |   |   |   |   |--- Temperature >  5568.50
|   |   |   |   |   |   |   |   |   |   |--- R <= 876.95
|   |   |   |   |   |   |   |   |   |   |   |--- class: 120000
|   |   |   |   |   |   |   |   |   |   |--- R >  876.95
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch o
f depth 9
|   |   |   |--- Temperature >  26256.50
|   |   |   |   |--- A_M <= -5.96
|   |   |   |   |   |--- Temperature <= 29289.00
|   |   |   |   |   |   |--- class: 849420
|   |   |   |   |   |--- Temperature >  29289.00
|   |   |   |   |   |   |--- Temperature <= 31664.00
|   |   |   |   |   |   |   |--- class: 834042
|   |   |   |   |   |   |--- Temperature >  31664.00
|   |   |   |   |   |   |   |--- Type <= 4.50
|   |   |   |   |   |   |   |   |--- Temperature <= 33525.00
|   |   |   |   |   |   |   |   |   |--- class: 240000
|   |   |   |   |   |   |   |   |--- Temperature >  33525.00
|   |   |   |   |   |   |   |   |   |--- Temperature <= 36875.00
|   |   |   |   |   |   |   |   |   |   |--- class: 220000
|   |   |   |   |   |   |   |   |   |--- Temperature >  36875.00
|   |   |   |   |   |   |   |   |   |   |--- class: 813000
|   |   |   |   |   |   |   |   |--- Type >  4.50
|   |   |   |   |   |   |   |   |   |--- A_M <= -10.39
|   |   |   |   |   |   |   |   |   |   |--- class: 648430
|   |   |   |   |   |   |   |   |   |--- A_M >  -10.39
```

```
|   |   |   |   |   |   |   |   |   |   |--- R <= 1569.50
|   |   |   |   |   |   |   |   |   |   |   |--- A_M <= -9.61
|   |   |   |   |   |   |   |   |   |   |   |   |--- class: 374830
|   |   |   |   |   |   |   |   |   |   |   |--- A_M >  -9.61
|   |   |   |   |   |   |   |   |   |   |   |   |--- class: 272830
|   |   |   |   |   |   |   |   |   |   |--- R >  1569.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: 294903
|   |   |   |   |   |--- A_M >  -5.96
|   |   |   |   |   |   |--- R <= 8.53
|   |   |   |   |   |   |   |--- R <= 6.13
|   |   |   |   |   |   |   |   |--- class: 188000
|   |   |   |   |   |   |   |--- R >  6.13
|   |   |   |   |   |   |   |   |--- R <= 6.27
|   |   |   |   |   |   |   |   |   |--- class: 173800
|   |   |   |   |   |   |   |   |--- R >  6.27
|   |   |   |   |   |   |   |   |   |--- R <= 6.35
|   |   |   |   |   |   |   |   |   |   |--- class: 28840
|   |   |   |   |   |   |   |   |   |--- R >  6.35
|   |   |   |   |   |   |   |   |   |   |--- A_M <= -4.57
|   |   |   |   |   |   |   |   |   |   |   |--- class: 198200
|   |   |   |   |   |   |   |   |   |   |--- A_M >  -4.57
|   |   |   |   |   |   |   |   |   |   |   |--- R <= 6.63
|   |   |   |   |   |   |   |   |   |   |   |   |--- class: 16790
|   |   |   |   |   |   |   |   |   |   |   |--- R >  6.63
|   |   |   |   |   |   |   |   |   |   |   |   |--- class: 202900
|   |   |   |   |   |   |--- R >  8.53
|   |   |   |   |   |   |   |--- Temperature <= 37554.00
|   |   |   |   |   |   |   |   |--- R <= 53.00
|   |   |   |   |   |   |   |   |   |--- class: 198000
|   |   |   |   |   |   |   |   |--- R >  53.00
|   |   |   |   |   |   |   |   |   |--- class: 352000
|   |   |   |   |   |   |   |--- Temperature >  37554.00
|   |   |   |   |   |   |   |   |--- class: 204000
|---- A_M >  4.57
|   |--- class: 0
```

In [14]:
```
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error, mean_squared_error,  media
```

In [15]:
```
forest_1 = RandomForestRegressor(n_estimators=5, oob_score=True, random_sta
forest_1.fit(X, Y)
```

Out[15]:
```
RandomForestRegressor(n_estimators=5, oob_score=True, random_state=10)
```

In [16]:
```
Y_predict = forest_1.predict(X_test)
print('Средняя абсолютная ошибка:',    mean_absolute_error(Y_test, Y_predict
print('Средняя квадратичная ошибка:', mean_squared_error(Y_test, Y_predict
print('Median absolute error:',        median_absolute_error(Y_test, Y_predi
print('Коэффициент детерминации:',     r2_score(Y_test, Y_predict))
```

```
Средняя абсолютная ошибка: 29120.574999999997
Средняя квадратичная ошибка: 5203187720.595
Median absolute error: 0.0
Коэффициент детерминации: 0.9230465292619281
```

In [17]:

```
plt.scatter(X_test.R, Y_test,    marker = 'o', label = 'Тестовая выборка')
plt.scatter(X_test.R, Y_predict, marker = '.', label = 'Предсказанные данны
plt.legend(loc = 'lower right')
plt.xlabel('R')
plt.ylabel('L')
plt.show()
```