**Московский государственный технический университет им. Н.Э. Баумана.**

Факультет «Информатика и системы управления»

Кафедра «Системы обработки информации и управления»

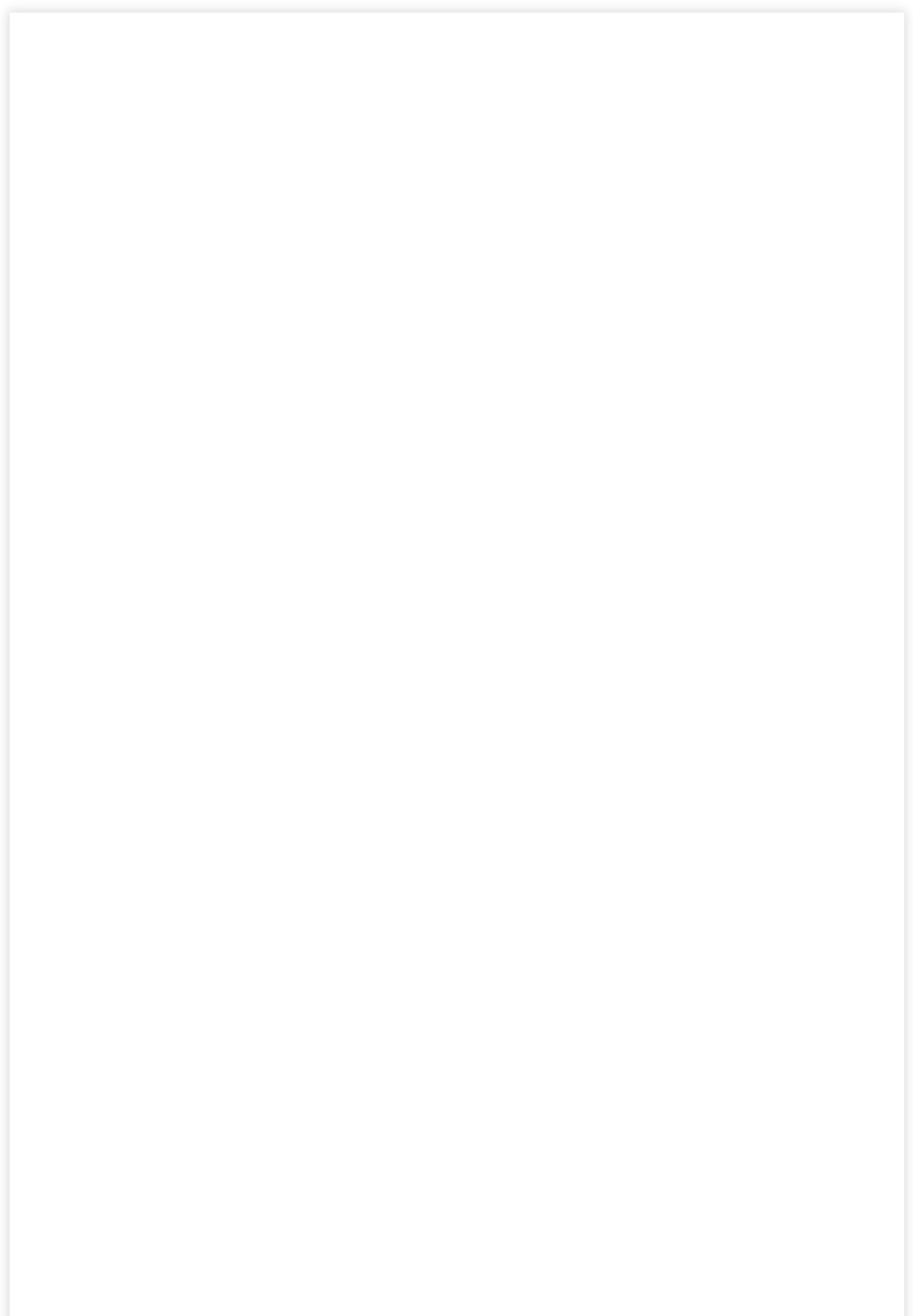
Курс «Технологии машинного обучения»

Отчет по лабораторной работе №6

Выполнил: Проверил:

Студент группы ИУ5-62Б преподаватель каф. ИУ5   
Грачев Я.А. Гапанюк Ю.Е.   
Подпись и дата: Подпись и дата:

Москва, 2020 г.



**Лабораторная работа №6**

**Ансамбли моделей машинного обучения**

**Цель лабораторной работы**

Изучение ансамблей моделей машинного обучения.

**Задание**

Выберите набор данных (датасет) для решения задачи классификации или регресии.

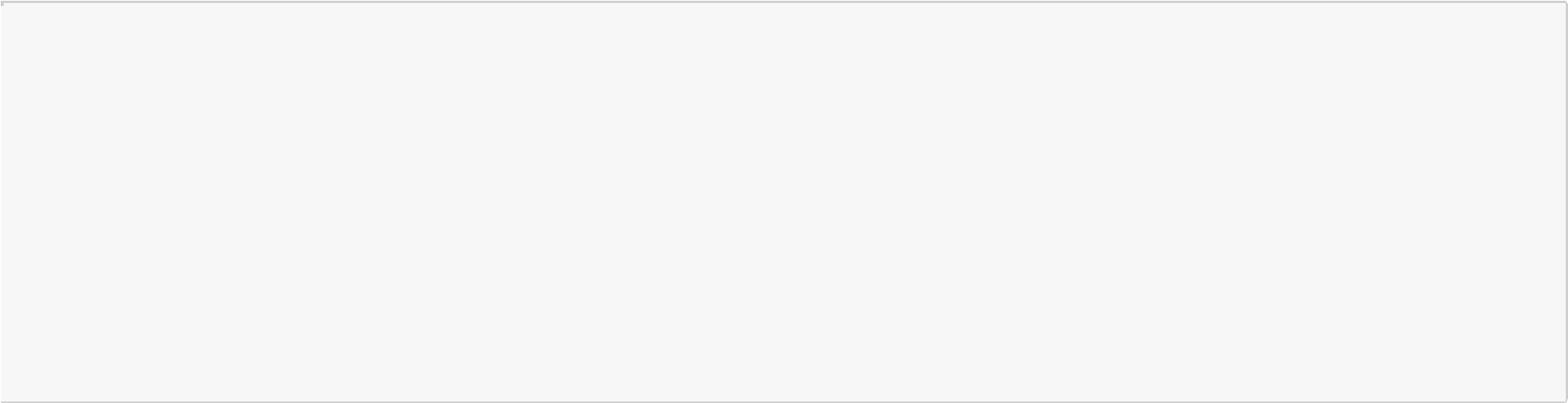
В случае необходимости проведите удаление или заполнение пропусков и кодирование категориальных признаков.

С использованием метода train\_test\_split разделите выборку на обучающую и тестовую.

Обучите две ансамблевые модели. Оцените качество моделей с помощью одной из подходящих для задачи метрик.

Сравните качество полученных моделей.

**Ход выполнения лабораторной работы**



In [1]:

**import**

**pandas**

**as**

**pd**

**import**

**seaborn**

**as**

**sns**

**import**

**numpy**

**as**

**np**

**import**

**matplotlib.pyplot**

**as**

**plt**

**from**

**sklearn.preprocessing**

**import**

MinMaxScaler

**from**

**sklearn.model\_selection**

**import**

train\_test\_split

**from**

**sklearn.metrics**

**import**

f1\_score

,

precision\_score

,

recall\_score

,

accuracy\_score

**from**

**sklearn.ensemble**

**import**

RandomForestClassifier

**from**

**sklearn.ensemble**

**import**

GradientBoostingClassifier

%

**matplotlib**

inline

sns

.

set

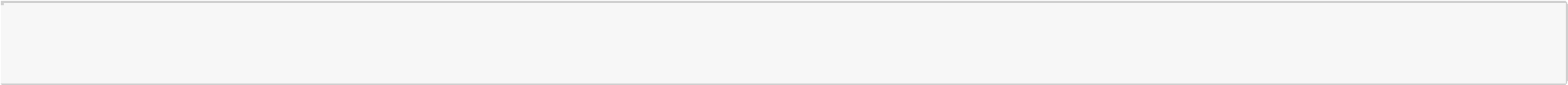
(

style

=

"ticks"

)



In [2]:

data

=

pd

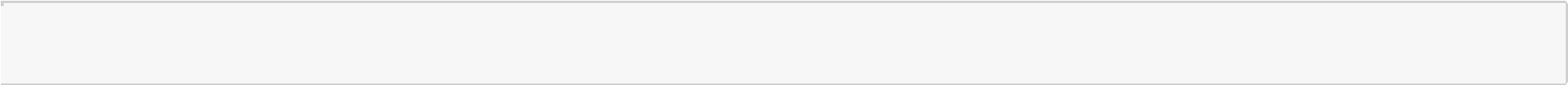
.

read\_csv

(

'pulsar\_stars.csv'

)



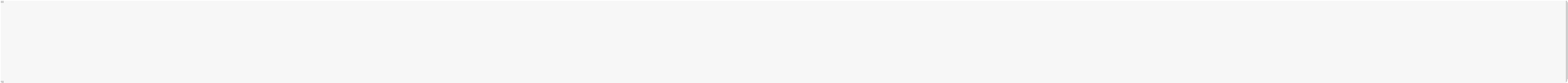
In [3]:

data

.

head

()



In [4]:

data

.

isnull

()

.

sum

()

Out[3]:

**Mean of the**

**integrated**

**profile**

**Standard**

**deviation of the**

**integrated**

**profile**

**Excess**

**kurtosis of the**

**integrated**

**profile**

**Skewness of**

**the**

**integrated**

**profile**

**Mean of**

**the DM-**

**SNR curve**

**Standard**

**deviation of the**

**DM-SNR curve**

**Excess**

**kurtosis of the**

**DM-SNR curve**

**Skewness of**

**the DM-SNR**

**curve**

**target\_class**

**0**

140.562500

55.683782

-0.234571

-0.699648

3.199833

19.110426

7.975532

74.242225

0

**1**

102.507812

58.882430

0.465318

-0.515088

1.677258

14.860146

10.576487

127.393580

0

**2**

103.015625

39.341649

0.323328

1.051164

3.121237

21.744669

7.735822

63.171909

0

**3**

136.750000

57.178449

-0.068415

-0.636238

3.642977

20.959280

6.896499

53.593661

0

**4**

88.726562

40.672225

0.600866

1.123492

1.178930

11.468720

14.269573

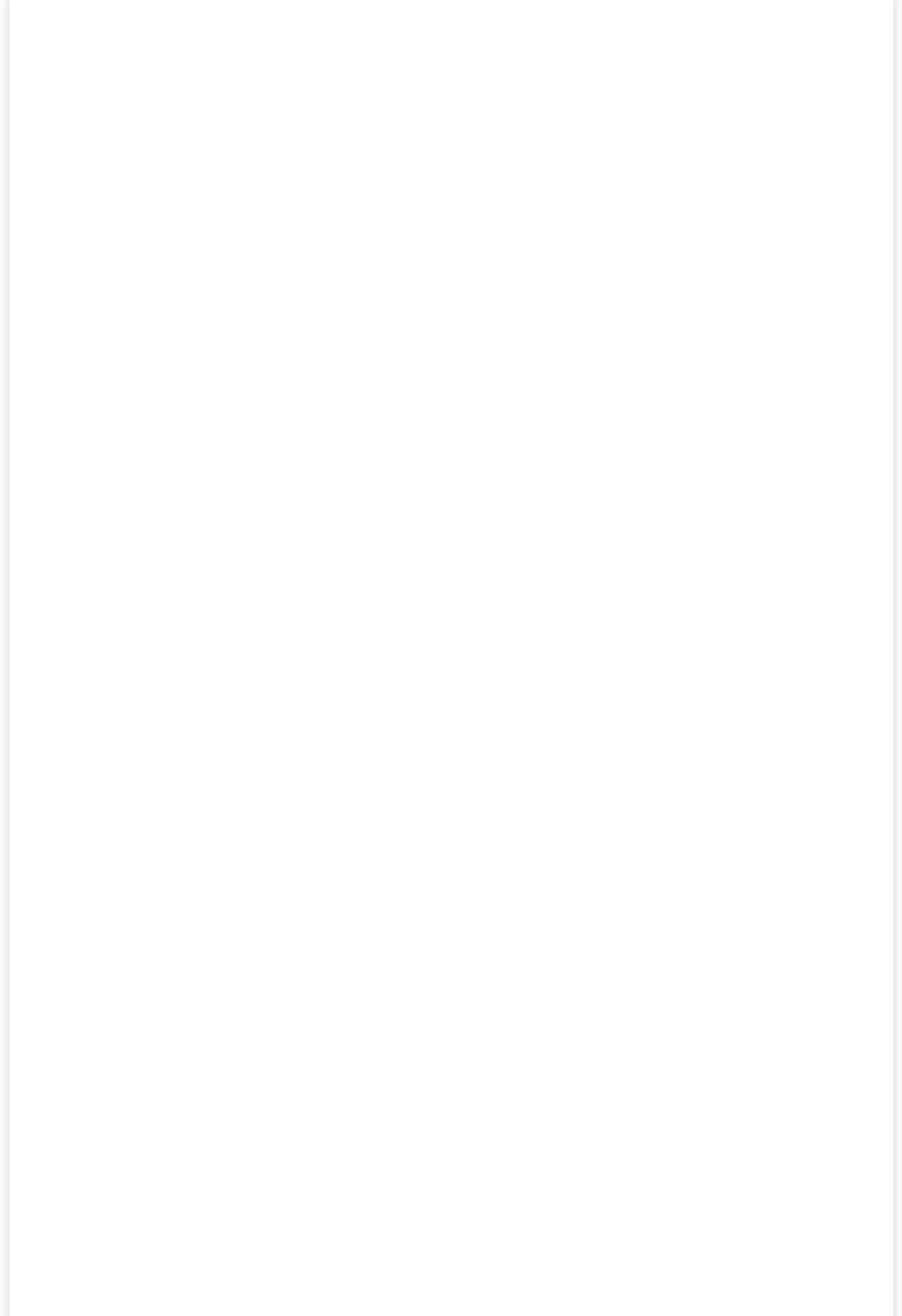
252.567306

0

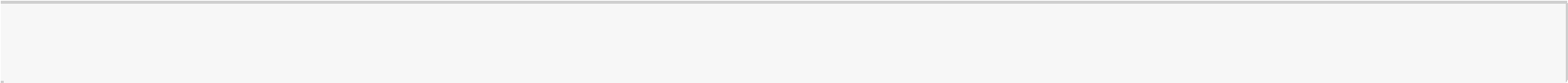
Out[4]:

Mean of the integrated profile 0





**Формирование обучающей и тестовой выборок**

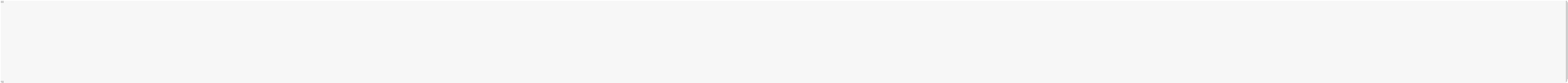


In [6]:

data

.

columns

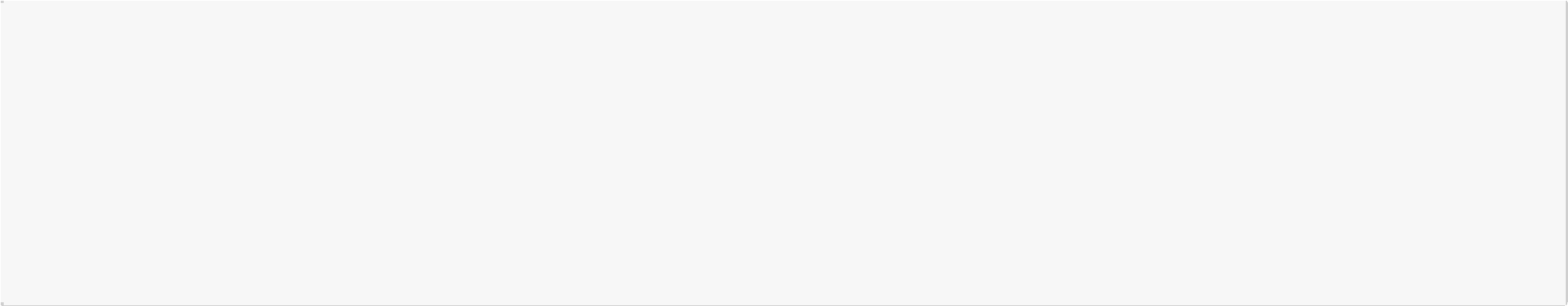


In [7]:

data

.

dtypes



In [8]:

*# Признаки для задачи классификации*

class\_cols

=

[

' Mean of the integrated profile'

,

' Standard deviation of the integrated profile'

,

' Excess kurtosis of the integrated profile'

,

' Skewness of the integrated profile'

,

' Mean of the DM-SNR curve'

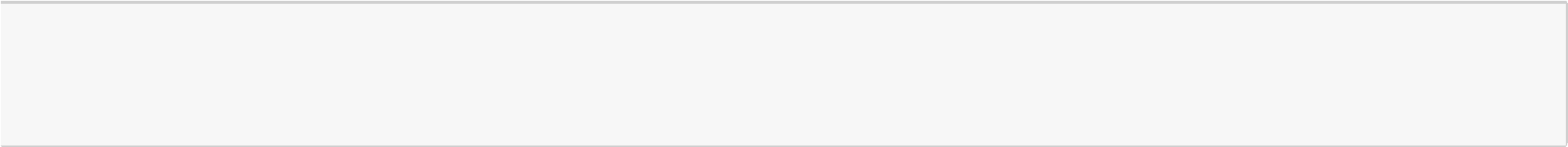
,

' Standard deviation of the DM-SNR curve'

,

' Excess kurtosis of the DM-SNR curve'

]



In [9]:

X

=

data

[

class\_cols

]

Y

=

data

[

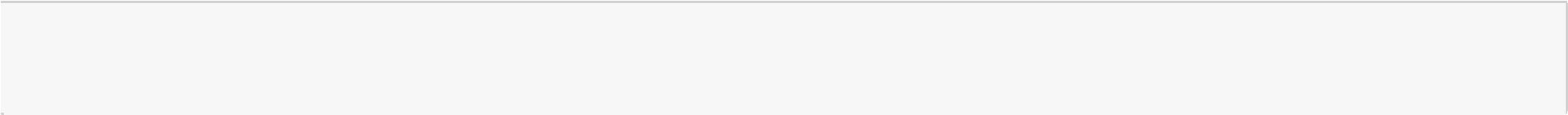
'target\_class'

]

X

.

shape



In [10]:

*# С использованием метода train\_test\_split разделим выборку на обучающую и тестовую*

X\_train

,

X\_test

,

Y\_train

,

Y\_test

=

train\_test\_split

(

X

,

Y

,

test\_size

=

0.25

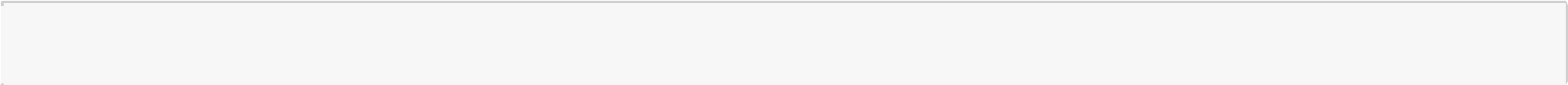
,

random\_state

=

1

)



In [11]:

X\_train

.

shape

,

X\_test

.

shape

,

Y\_train

.

shape

,

Y\_test

.

shape

Out[6]:

Index([' Mean of the integrated profile',

' Standard deviation of the integrated profile',

' Excess kurtosis of the integrated profile',

' Skewness of the integrated profile', ' Mean of the DM-SNR curve',

' Standard deviation of the DM-SNR curve',

' Excess kurtosis of the DM-SNR curve', ' Skewness of the DM-SNR curve',

'target\_class'],

dtype='object')

Out[7]:

Mean of the integrated profile float64

Standard deviation of the integrated profile float64

Excess kurtosis of the integrated profile float64

Skewness of the integrated profile float64

Mean of the DM-SNR curve float64

Standard deviation of the DM-SNR curve float64

Excess kurtosis of the DM-SNR curve float64

Skewness of the DM-SNR curve float64

target\_class int64

dtype: object

Out[9]:

(17898, 7)

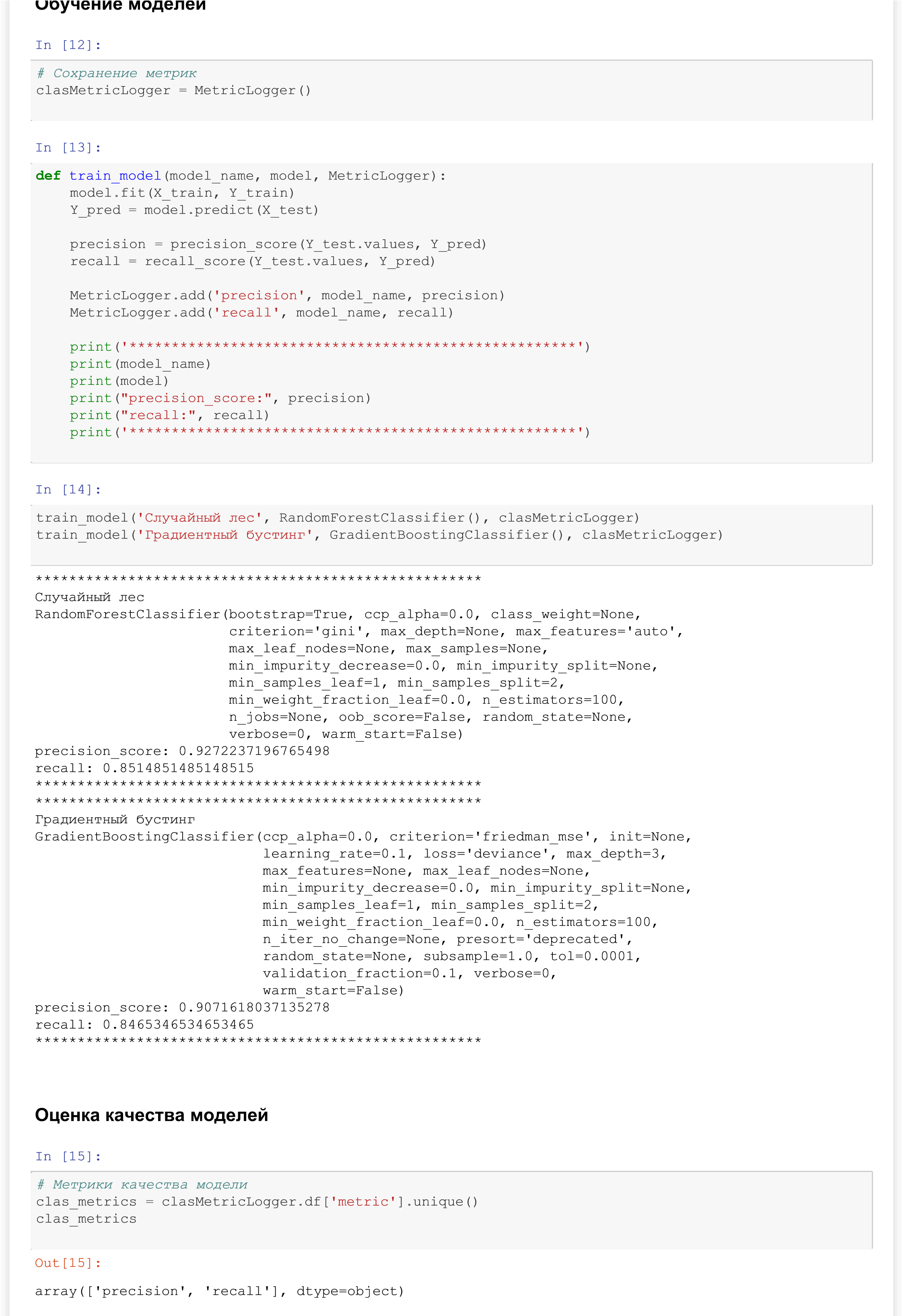
Out[11]:

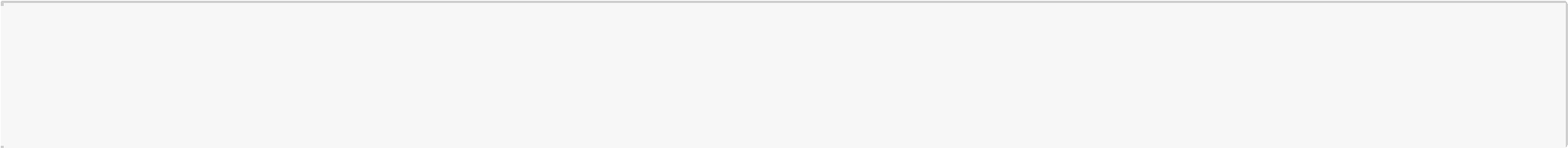
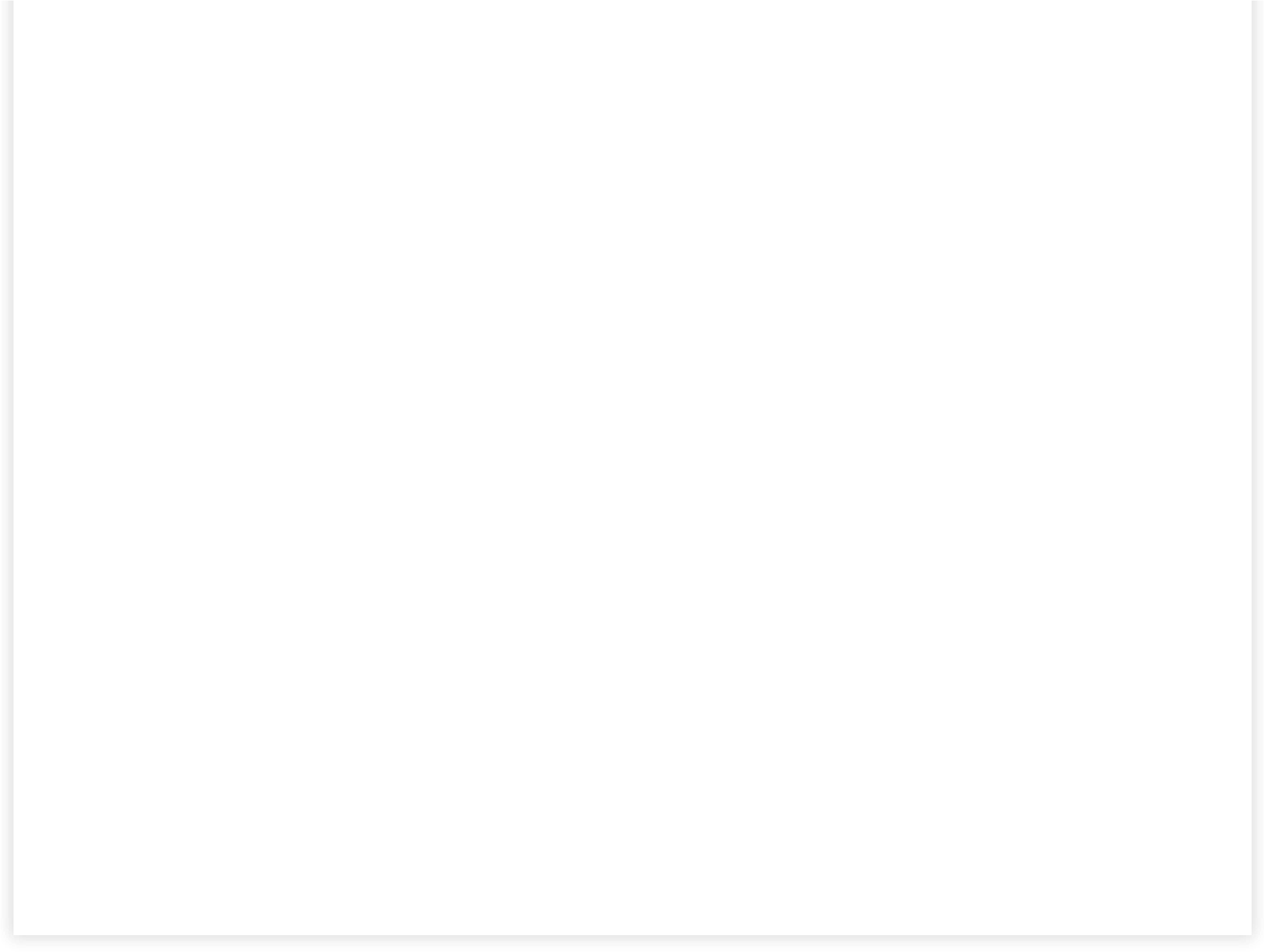
((13423, 7)

, (4475, 7), (13423,), (4475,

))

**Обучение моделей**





In [16]:

*# Построим графики метрик качества модели*

**for**

metric

**in**

clas\_metrics

:

clasMetricLogger

.

plot

(

'Метрика: '

+

metric

,

metric

,

figsize

=

(

5

,

3

))

Таким образом, на основании обеих метрик лучшей моделью оказался случайный лес.

