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.

Метод 1: Линейная/логистическая регрессия

Метод 2: Градиентный бустинг

Набор данных: https://www.kaggle.com/datasets/fedesoriano/company-bankruptcy-prediction

Данные были получены из Тайваньского экономического журнала за период с 1999 по 2009 год. Банкротство компании было определено на основании правил ведения бизнеса Тайваньской фондовой биржи.

Столбцы:

Y - Банкрот?: Ярлык класса

X1 - ROA(C) до вычета процентов и амортизации до вычета процентов: Рентабельность общих активов(C)

X2 - ROA(A) до уплаты процентов и % после налогообложения: Рентабельность общих активов(A)

X3 - ROA (B) до вычета процентов и амортизации после налогообложения: Рентабельность всех активов(B)

X4 – Валовая прибыль от операционной деятельности: Валовая прибыль/чистая выручка от продаж

X5 – Валовая прибыль от реализованных продаж: Реализованная валовая прибыль/чистая выручка от продаж

X6 – Норма операционной прибыли: Операционная прибыль/ Чистая выручка

X7 – Чистая процентная ставка до налогообложения: Доход/чистая выручка

X8 – Чистая процентная ставка после уплаты налогов: Чистая прибыль/Чистая выручка

X9 – Непромышленные доходы и расходы/выручка: Коэффициент чистой внереализационной прибыли

X10 — Непрерывная процентная ставка (после налогообложения): Чистая прибыль — исключая прибыль или убыток от выбытия / Чистая выручка

X11 — Ставка операционных расходов: Операционные расходы / Чистая выручка

X12 — Ставка расходов на исследования и разработки: (Расходы на исследования и разработки) / Чистая выручка

X13 - Скорость движения денежных средств: Денежный поток от операционных/текущих обязательств

X14 - Процентная ставка по процентным долгам: Процентные долги/капитал

X15 - Ставка налога (A): Эффективная налоговая ставка

X16 - Чистая стоимость одной акции (B): Балансовая стоимость На акцию(B)

X17 - Чистая стоимость на акцию (A): Балансовая стоимость на акцию(A)

X18 - Чистая стоимость на акцию (C): Балансовая стоимость на акцию(C)

X19 - Постоянная прибыль на акцию за последние четыре сезона: прибыль на акцию -Чистый доход

X20 - Денежный поток на акцию

X21 — Доход на акцию (в юанях): Продажи на акцию

X22 — Операционная прибыль на акцию (в юанях): Операционный доход на акцию

X23 — Чистая прибыль на акцию до налогообложения (в юанях): Доход на акцию до налогообложения X24 — Валовой реализованный доход Темп роста прибыли

X25 - Темп роста операционной прибыли: Рост операционной прибыли

X26 - Темп роста чистой прибыли после уплаты налогов: Рост чистой прибыли

X27 - Темп роста обычной чистой прибыли: Продолжающийся рост операционной прибыли после налогообложения

X28 - Темп постоянного роста чистой прибыли: Чистая прибыль - Исключая рост прибыли или убытков от выбытия

X29 - Темп роста общих активов: Рост совокупных активов

X30 - Темп роста чистой стоимости: Рост совокупных активов X31 - Коэффициент темпов роста общей доходности активов: Прибыль к совокупным приростам активов

X32 - Реинвестирование денежных средств, %: Коэффициент реинвестирования денежных средств

X33 - Коэффициент текущей ликвидности

X34 - Коэффициент быстрой ликвидности: тест на кислотность

X35 - Коэффициент процентных расходов: процентные расходы/общий доход

X36 - Общий долг/Общий собственный капитал: Коэффициент общих обязательств/капитала

X37 - Коэффициент долга, % : Обязательства/Общие активы

X38 - Чистая стоимость/Активы: Собственный капитал/Общие активы

X39 - Коэффициент пригодности долгосрочного фонда (A): (Долгосрочные обязательства + Собственный капитал)/Основные активы

X40 - Зависимость от заимствования: Стоимость процентного долга

X41 – Условные обязательства/Собственный капитал: Условные обязательства/Капитал

X42 – Операционная прибыль/Оплаченный капитал: Операционный доход/Капитал

X43 – Чистая прибыль до налогообложения/Оплаченный капитал: Доход/капитал до налогообложения

X44 — Товарно-материальные запасы и дебиторская задолженность/Чистая стоимость: (Запасы+Дебиторская задолженность)/Собственный капитал

X45 — Общий оборот активов

X46 — Оборачиваемость дебиторской задолженности

X47 — Среднее количество дней сбора: Дни непогашенной дебиторской задолженности

X48 — Коэффициент оборачиваемости запасов (в разах)

X49 — Оборачиваемость основных средств Частота

X50 - Коэффициент оборота чистой стоимости (раз): Оборачиваемость собственного капитала

X51 - Выручка на человека: Продажи на одного работника

X52 - Операционная прибыль на человека: Операционный доход на одного работника

X53 - Коэффициент распределения на человека: Основные средства на одного работника

X54 - Общий оборотный капитал Активы

X55 - Быстрые активы/Итого активы

X56 - Оборотные активы/Итого активы

X57 - Денежные средства/Итого активы

X58 - Оборотные активы/текущие обязательства

X59 - Денежные средства/текущие обязательства

X60 - Текущие обязательства по активам

X61 - Оборотные средства по обязательствам

X62 - Запасы/оборотные средства

X63 - Запасы/текущие обязательства

X64 - Текущие обязательства/обязательства

X65 - Оборотные средства/капитал

X66 - Текущие обязательства/капитал

X67 - Долгосрочные обязательства по текущим активам

X68 - Нераспределенная прибыль к общей сумме активов

X69 - Общая прибыль/общая сумма расходов

X70 - Общая сумма расходов/активов

X71 - Коэффициент оборачиваемости текущих активов: Текущие активы к продажам

X72 - Быстрая оборачиваемость активов Отношение оборотных средств к продажам

X73 - Коэффициент оборачиваемости оборотного капитала: оборотный капитал к продажам

X74 - Норма денежного оборота: Денежные средства к продажам

X75 - Денежные потоки к продажам

X76 - Основные средства к активам

X77 - Текущие обязательства к обязательствам

X78 - Текущие обязательства к капиталу

X79 - Капитал к долгосрочным обязательствам

X80 - Денежные потоки к общим активам

X81 - Денежный поток к обязательствам

X82 - Финансовый директор к активам

X83 - Денежный поток к собственному капиталу

X84 - Текущие обязательства к оборотным активам

X85 - Флаг пассивов-активов: 1, если общая сумма обязательств превышает общую сумму активов, 0 в противном случае

X86 - чистая прибыль к общей сумме активов

X87 - общая сумма активов к цене ВНП

X88 - Интервал без кредита

X89 - Валовая прибыль к продажам

X90 - Чистая прибыль к акционерному капиталу

X91 - Обязательства к капиталу

X92 — Степень финансового рычага (DFL)

X93 — Коэффициент покрытия процентов (процентные расходы к EBIT)

X94 — Флаг чистой прибыли: 1, если чистая прибыль отрицательная за последние два года, 0 в противном случае

X95 — отношение капитала к обязательствам

Ход работы:

Подключаем все необходимые библиотеки:

```
import pandas as pd
import seaborn as sb
```

```
import numpy as np
import matplotlib
import matplotlib_inline
import matplotlib.pyplot as plt
from IPython.display import Image
from io import StringIO
import graphviz
import pydotplus
from sklearn.model_selection import train_test_split
%matplotlib inline
sb.set(style="ticks")
from IPython.display import set_matplotlib_formats
#matplotlib_inline.backend_inline.set_matplotlib_formats("retina")
```

Подключаем датасет

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

Размер набора данных

```
data.shape
```

```
(6819, 96)
```

Типы колонок

```
data.dtypes
```

```
Bankrupt?                                                   int64
 ROA(C) before interest and depreciation before interest    float64
 ROA(A) before interest and % after tax                     float64
 ROA(B) before interest and depreciation after tax          float64
 Operating Gross Margin                                     float64
                                                             ...
 Liability to Equity                                        float64
 Degree of Financial Leverage (DFL)                         float64
 Interest Coverage Ratio (Interest expense to EBIT)         float64
 Net Income Flag                                             int64
 Equity to Liability                                        float64
Length: 96, dtype: object
```

Проверяем есть ли пропущенные значения

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

```
Bankrupt?                                                   0
 ROA(C) before interest and depreciation before interest    0
 ROA(A) before interest and % after tax                     0
 ROA(B) before interest and depreciation after tax          0
 Operating Gross Margin                                     0
                                                           ..
 Liability to Equity                                        0
```

```
 Degree of Financial Leverage (DFL)                          0
 Interest Coverage Ratio (Interest expense to EBIT)          0
 Net Income Flag                                             0
 Equity to Liability                                         0
Length: 96, dtype: int64
```

В наборе нет пропусков, следовательно не нужно их обрабатывать.

Возьмем для анализа первые 2000 строк набора данных

```
data_2t = data.head(2000)
```

Первые 5 строк датасета

```
data_2t.head(5)
```

```
   Bankrupt?   ROA(C) before interest and depreciation before interest
\
0          1                                                 0.370594

1          1                                                 0.464291

2          1                                                 0.426071

3          1                                                 0.399844

4          1                                                 0.465022


    ROA(A) before interest and % after tax  \
0                                 0.424389
1                                 0.538214
2                                 0.499019
3                                 0.451265
4                                 0.538432

    ROA(B) before interest and depreciation after tax  \
0                                           0.405750
1                                           0.516730
2                                           0.472295
3                                           0.457733
4                                           0.522298

    Operating Gross Margin   Realized Sales Gross Margin  \
0                 0.601457                      0.601457
1                 0.610235                      0.610235
2                 0.601450                      0.601364
3                 0.583541                      0.583541
4                 0.598783                      0.598783

    Operating Profit Rate   Pre-tax net Interest Rate  \
```

```
0                0.998969                    0.796887
1                0.998946                    0.797380
2                0.998857                    0.796403
3                0.998700                    0.796967
4                0.998973                    0.797366

    After-tax net Interest Rate   Non-industry income and
expenditure/revenue  \
0                     0.808809
0.302646
1                     0.809301
0.303556
2                     0.808388
0.302035
3                     0.808966
0.303350
4                     0.809304
0.303475

   ...   Net Income to Total Assets   Total assets to GNP price  \
0  ...                     0.716845                    0.009219
1  ...                     0.795297                    0.008323
2  ...                     0.774670                    0.040003
3  ...                     0.739555                    0.003252
4  ...                     0.795016                    0.003878

    No-credit Interval   Gross Profit to Sales  \
0             0.622879                0.601453
1             0.623652                0.610237
2             0.623841                0.601449
3             0.622929                0.583538
4             0.623521                0.598782

    Net Income to Stockholder's Equity   Liability to Equity  \
0                             0.827890              0.290202
1                             0.839969              0.283846
2                             0.836774              0.290189
3                             0.834697              0.281721
4                             0.839973              0.278514

    Degree of Financial Leverage (DFL)  \
0                             0.026601
1                             0.264577
2                             0.026555
3                             0.026697
4                             0.024752

    Interest Coverage Ratio (Interest expense to EBIT)   Net Income
Flag  \
0                                           0.564050
```

```
1
1                                                0.570175
1
2                                                0.563706
1
3                                                0.564663
1
4                                                0.575617
1

    Equity to Liability
0              0.016469
1              0.020794
2              0.016474
3              0.023982
4              0.035490

[5 rows x 96 columns]
```

```
from IPython.display import set_matplotlib_formats
#matplotlib.use('nbagg')
#print(matplotlib.get_backend())
#matplotlib_inline.backend_inline.set_matplotlib_formats("retina")
pd.set_option("display.width", 70)
```

Нет категориальных значений, значит ненужно кодировать категориальных признаков.

Масштабирование данных

```
from sklearn.preprocessing import MinMaxScaler
scl = MinMaxScaler ()
scl_data = scl.fit_transform(data_2t)
data_scalled = data_2t.copy()
data_scalled[data_scalled.columns] = scl_data
data_scalled
```

```
      Bankrupt?  \
0           1.0
1           1.0
2           1.0
3           1.0
4           1.0
...         ...
1995        0.0
1996        0.0
1997        0.0
1998        0.0
1999        0.0

       ROA(C) before interest and depreciation before interest  \
```

0	0.505452
1	0.633245
2	0.581117
3	0.545346
4	0.634242
...	...
1995	0.652128
1996	0.551995
1997	0.651662
1998	0.697207
1999	0.710306

	ROA(A) before interest and % after tax
0	0.564662
1	0.716109
2	0.663959
3	0.600421
4	0.716400
...	...
1995	0.661275
1996	0.606804
1997	0.713353
1998	0.753318
1999	0.747008

	ROA(B) before interest and depreciation after tax
0	0.552325
1	0.703396
2	0.642909
3	0.623087
4	0.710975
...	...
1995	0.700554
1996	0.623160
1997	0.717097
1998	0.753170
1999	0.770223

	Operating Gross Margin	Realized Sales Gross Margin
0	0.874827	0.874827
1	0.892077	0.892077
2	0.874812	0.874642
3	0.839617	0.839617
4	0.869572	0.869572
...	...	...
1995	0.878311	0.878311
1996	0.869374	0.869317
1997	0.882857	0.880817
1998	0.884103	0.884967
1999	0.879543	0.879543

```
      Operating Profit Rate   Pre-tax net Interest Rate  \
0                  0.981222                    0.934741
1                  0.980329                    0.945235
2                  0.976925                    0.924444
3                  0.970871                    0.936440
4                  0.981372                    0.944934
...                     ...                         ...
1995               0.977453                    0.941628
1996               0.981559                    0.941316
1997               0.982464                    0.945449
1998               0.982455                    0.950381
1999               0.982748                    0.948276

      After-tax net Interest Rate  \
0                        0.955346
1                        0.965279
2                        0.946819
3                        0.958505
4                        0.965340
...                           ...
1995                     0.961634
1996                     0.962266
1997                     0.965524
1998                     0.968732
1999                     0.967772

      Non-industry income and expenditure/revenue  ...  \
0                                        0.827681  ...
1                                        0.851899  ...
2                                        0.811414  ...
3                                        0.846393  ...
4                                        0.849732  ...
...                                           ...  ...
1995                                     0.848188  ...
1996                                     0.841556  ...
1997                                     0.849275  ...
1998                                     0.860062  ...
1999                                     0.855039  ...

      Net Income to Total Assets   Total assets to GNP price  \
0                       0.800539                9.388432e-13
1                       0.888151                8.475867e-13
2                       0.865115                4.073610e-12
3                       0.825900                3.312093e-13
4                       0.887837                3.948639e-13
...                          ...                         ...
1995                    0.861755                1.276465e-13
1996                    0.829739                1.024685e-13
1997                    0.890815                5.799301e-14
```

1998	0.909790	1.456608e-13
1999	0.905668	7.207471e-14

	No-credit Interval	Gross Profit to Sales	\
0	0.362237	0.874821	
1	0.363543	0.892083	
2	0.363864	0.874814	
3	0.362321	0.839613	
4	0.363322	0.869571	
...	...	...	
1995	0.364198	0.878312	
1996	0.364453	0.869374	
1997	0.363918	0.882859	
1998	0.363477	0.884099	
1999	0.361825	0.879540	

	Net Income to Stockholder's Equity	Liability to Equity	\
0	0.827890	0.389349	
1	0.839969	0.380821	
2	0.836774	0.389331	
3	0.834697	0.377971	
4	0.839973	0.373667	
...	...	...	
1995	0.837870	0.376931	
1996	0.835980	0.374476	
1997	0.840176	0.374805	
1998	0.841449	0.374193	
1999	0.841382	0.376910	

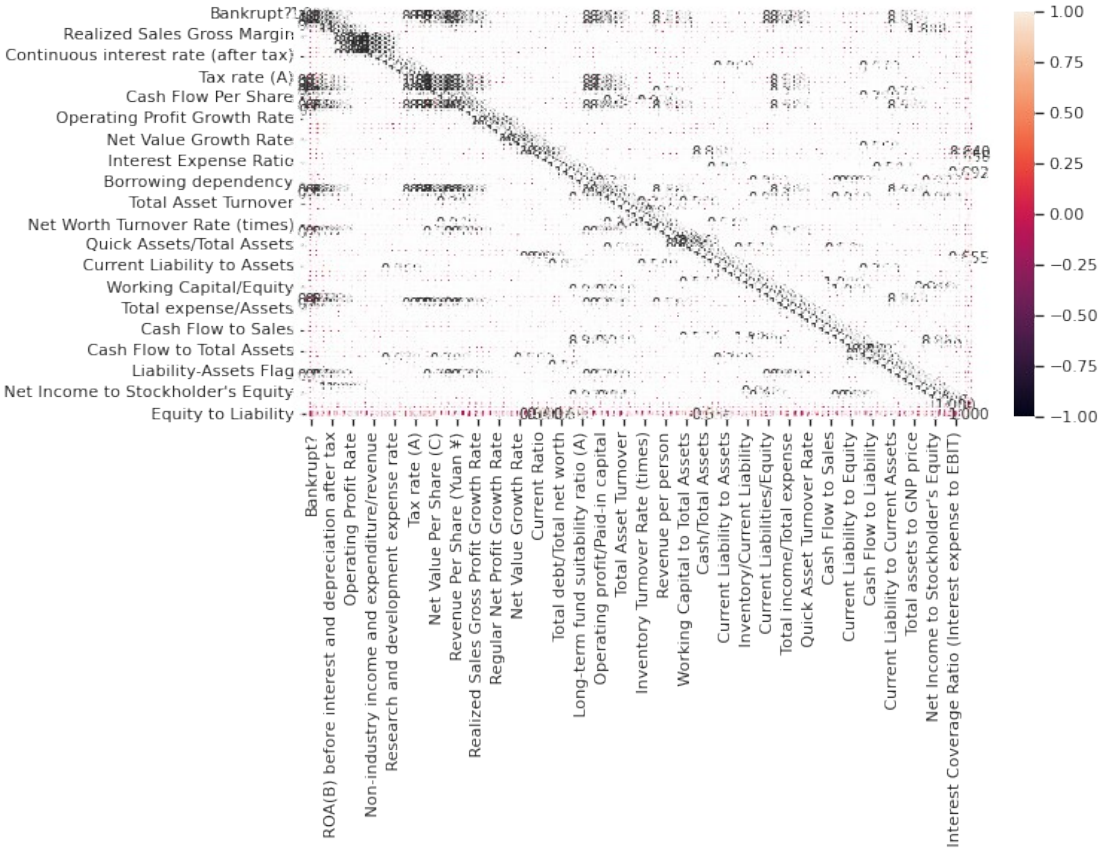
	Degree of Financial Leverage (DFL)	\
0	0.025832	
1	0.263996	
2	0.025786	
3	0.025928	
4	0.023981	
...	...	
1995	0.025880	
1996	0.025943	
1997	0.026758	
1998	0.026121	
1999	0.026160	

	Interest Coverage Ratio (Interest expense to EBIT)	\
0	0.564050	
1	0.570175	
2	0.563706	
3	0.564663	
4	0.575617	
...	...	
1995	0.564377	

```
1996                                                0.564753
1997                                                0.567226
1998                                                0.565590
1999                                                0.565737

      Net Income Flag   Equity to Liability
0                  0.0              0.009528
1                  0.0              0.015009
2                  0.0              0.009534
3                  0.0              0.019048
4                  0.0              0.033631
...                ...                   ...
1995               0.0              0.021203
1996               0.0              0.029281
1997               0.0              0.027831
1998               0.0              0.030664
1999               0.0              0.021254

[2000 rows x 96 columns]
```

```
ig, ax = plt.subplots(figsize=(10,5))
sb.heatmap(data_scalled.corr(method='pearson'),ax=ax, annot=True,
fmt='.3f')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fdf2cb89e90>
```




```
#data_scalled.dtypes
X = data_scalled.drop(columns=' ROA(C) before interest and
depreciation before interest')
Y = data_scalled[' ROA(C) before interest and depreciation before
interest']

X.head()
```

```
   Bankrupt?   ROA(A) before interest and % after tax  \
0        1.0                                 0.564662
1        1.0                                 0.716109
2        1.0                                 0.663959
3        1.0                                 0.600421
4        1.0                                 0.716400

    ROA(B) before interest and depreciation after tax  \
0                                           0.552325
1                                           0.703396
2                                           0.642909
3                                           0.623087
4                                           0.710975

    Operating Gross Margin   Realized Sales Gross Margin  \
0                 0.874827                      0.874827
```

```
1                0.892077                     0.892077
2                0.874812                     0.874642
3                0.839617                     0.839617
4                0.869572                     0.869572

   Operating Profit Rate   Pre-tax net Interest Rate  \
0               0.981222                    0.934741
1               0.980329                    0.945235
2               0.976925                    0.924444
3               0.970871                    0.936440
4               0.981372                    0.944934

   After-tax net Interest Rate  \
0                      0.955346
1                      0.965279
2                      0.946819
3                      0.958505
4                      0.965340

   Non-industry income and expenditure/revenue  \
0                                     0.827681
1                                     0.851899
2                                     0.811414
3                                     0.846393
4                                     0.849732

   Continuous interest rate (after tax)  ...  \
0                              0.949485  ...
1                              0.960267  ...
2                              0.934983  ...
3                              0.954784  ...
4                              0.961179  ...

   Net Income to Total Assets   Total assets to GNP price  \
0                    0.800539                9.388432e-13
1                    0.888151                8.475867e-13
2                    0.865115                4.073610e-12
3                    0.825900                3.312093e-13
4                    0.887837                3.948639e-13

   No-credit Interval   Gross Profit to Sales  \
0            0.362237                0.874821
1            0.363543                0.892083
2            0.363864                0.874814
3            0.362321                0.839613
4            0.363322                0.869571

   Net Income to Stockholder's Equity   Liability to Equity  \
0                            0.827890              0.389349
```

```
1                            0.839969             0.380821
2                            0.836774             0.389331
3                            0.834697             0.377971
4                            0.839973             0.373667

    Degree of Financial Leverage (DFL)  \
0                            0.025832
1                            0.263996
2                            0.025786
3                            0.025928
4                            0.023981

    Interest Coverage Ratio (Interest expense to EBIT)  \
0                                           0.564050
1                                           0.570175
2                                           0.563706
3                                           0.564663
4                                           0.575617

    Net Income Flag   Equity to Liability
0               0.0              0.009528
1               0.0              0.015009
2               0.0              0.009534
3               0.0              0.019048
4               0.0              0.033631

[5 rows x 95 columns]
```

```
Y.head()
```

```
0    0.505452
1    0.633245
2    0.581117
3    0.545346
4    0.634242
Name:  ROA(C) before interest and depreciation before interest, dtype:
float64
```

```
X_train,  X_test,  Y_train,  Y_test = train_test_split(X,  Y,
random_state = 2022, test_size = 0.1)
```

```
X_train.head()
```

```
      Bankrupt?   ROA(A) before interest and % after tax  \
1964        0.0                                  0.666933
1510        0.0                                  0.762748
228         0.0                                  0.721622
1189        0.0                                  0.739900
1889        0.0                                  0.673243

       ROA(B) before interest and depreciation after tax  \
```

1964		0.677015
1510		0.762717
228		0.734587
1189		0.741801
1889		0.668051

	Operating Gross Margin	Realized Sales Gross Margin	\
1964	0.867561	0.867901	
1510	0.882460	0.882460	
228	0.872263	0.872263	
1189	0.871427	0.871427	
1889	0.872872	0.872872	

	Operating Profit Rate	Pre-tax net Interest Rate	\
1964	0.980312	0.943662	
1510	0.982868	0.946562	
228	0.982580	0.945691	
1189	0.982837	0.945956	
1889	0.982109	0.944286	

	After-tax net Interest Rate	\
1964	0.964190	
1510	0.966593	
228	0.965832	
1189	0.965977	
1889	0.964645	

	Non-industry income and expenditure/revenue	\
1964	0.848497	
1510	0.851121	
228	0.849635	
1189	0.849841	
1889	0.847255	

	Continuous interest rate (after tax)	...	\
1964	0.960128	...	
1510	0.962590	...	
228	0.961916	...	
1189	0.962024	...	
1889	0.960642	...	

	Net Income to Total Assets	Total assets to GNP price	\
1964	0.868427	2.152637e-13	
1510	0.906061	8.215297e-14	
228	0.891599	4.496606e-13	
1189	0.896105	3.326637e-13	
1889	0.871126	4.043038e-14	

	No-credit Interval	Gross Profit to Sales	\

```
1964            0.364839               0.867559
1510            0.363946               0.882461
228             0.373931               0.872257
1189            0.361879               0.871432
1889            0.363506               0.872868

      Net Income to Stockholder's Equity   Liability to Equity  \
1964                            0.838824             0.372491
1510                            0.841345             0.376049
228                             0.840347             0.383954
1189                            0.840690             0.379043
1889                            0.838681             0.376182

      Degree of Financial Leverage (DFL)  \
1964                            0.025986
1510                            0.026510
228                             0.028023
1189                            0.027696
1889                            0.025959

      Interest Coverage Ratio (Interest expense to EBIT)  \
1964                                           0.564978
1510                                           0.566749
228                                            0.568455
1189                                           0.568246
1889                                           0.564838

      Net Income Flag   Equity to Liability
1964              0.0              0.043131
1510              0.0              0.023499
228               0.0              0.012291
1189              0.0              0.017271
1889              0.0              0.023118

[5 rows x 95 columns]

X_test.head()

      Bankrupt?   ROA(A) before interest and % after tax  \
1018        0.0                                  0.691231
1295        0.0                                  0.755639
643         0.0                                  0.744614
1842        0.0                                  0.722057
1669        0.0                                  0.762167

      ROA(B) before interest and depreciation after tax  \
1018                                           0.740490
1295                                           0.755502
643                                            0.776126
1842                                           0.725404
```

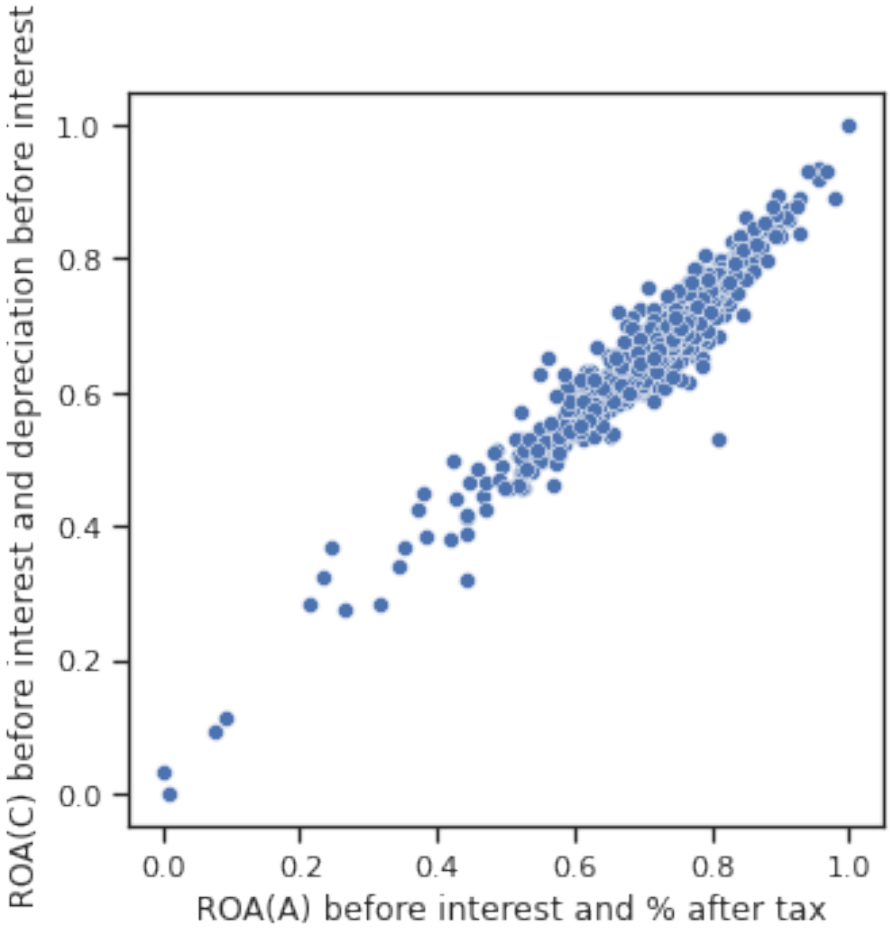
```
1669                                          0.775179

      Operating Gross Margin   Realized Sales Gross Margin  \
1018                0.911240                      0.907996
1295                0.873580                      0.873580
643                 0.895519                      0.895519
1842                0.877602                      0.877602
1669                0.916735                      0.916735

      Operating Profit Rate   Pre-tax net Interest Rate  \
1018               0.984657                    0.944793
1295               0.983494                    0.947260
643                0.984089                    0.947007
1842               0.982517                    0.945252
1669               0.986392                    0.949301

      After-tax net Interest Rate  \
1018                      0.964245
1295                      0.966832
643                       0.966885
1842                      0.965569
1669                      0.973196

      Non-industry income and expenditure/revenue  \
1018                                     0.844663
1295                                     0.851732
643                                      0.850324
1842                                     0.848767
1669                                     0.851996

      Continuous interest rate (after tax)  ...  \
1018                              0.964132  ...
1295                              0.963089  ...
643                               0.963246  ...
1842                              0.961661  ...
1669                              0.969847  ...

      Net Income to Total Assets   Total assets to GNP price  \
1018                    0.881355                2.815564e-14
1295                    0.910509                2.662760e-13
643                     0.905470                5.024619e-14
1842                    0.890484                2.226269e-13
1669                    0.913754                1.926183e-13

      No-credit Interval   Gross Profit to Sales  \
1018            0.363907                0.911235
1295            0.364176                0.873578
643             0.364205                0.895521
1842            0.363847                0.877608
```

```
1669             0.363471                0.916740

      Net Income to Stockholder's Equity   Liability to Equity  \
1018                            0.839582              0.372812
1295                            0.841734              0.376552
643                             0.840945              0.371431
1842                            0.840191              0.379881
1669                            0.841378              0.371266

      Degree of Financial Leverage (DFL)  \
1018                            0.025925
1295                            0.026067
643                             0.026152
1842                            0.065028
1669                            0.026064

      Interest Coverage Ratio (Interest expense to EBIT)  \
1018                                           0.564649
1295                                           0.565365
643                                            0.565708
1842                                           0.570064
1669                                           0.565348

      Net Income Flag   Equity to Liability
1018              0.0              0.040017
1295              0.0              0.022128
643               0.0              0.058259
1842              0.0              0.016117
1669              0.0              0.061687

[5 rows x 95 columns]

Y_train.head()

1964    0.609309
1510    0.696077
228     0.670545
1189    0.669947
1889    0.598670
Name:  ROA(C) before interest and depreciation before interest, dtype:
float64

Y_test.head()

1018    0.679388
1295    0.692553
643     0.701862
1842    0.657646
1669    0.678590
Name:  ROA(C) before interest and depreciation before interest, dtype:
float64
```

Линейная регрессия

```
from sklearn.linear_model import LinearRegression
from sklearn.datasets import make_blobs
from sklearn.metrics import mean_absolute_error, mean_squared_error

fig, ax = plt.subplots(figsize=(5,5))
sb.scatterplot(ax=ax, x=X[' ROA(A) before interest and % after tax'],
y=Y)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fdf23dff610>
```

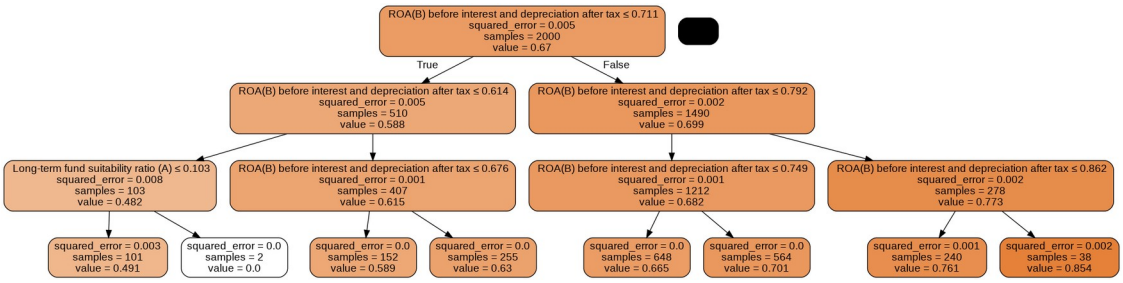



```
reg1 = LinearRegression().fit(X, Y)

Y_pred_1 = reg1.predict(X_test)

mean_absolute_error(Y_test, Y_pred_1), mean_squared_error(Y_test,
Y_pred_1)
```

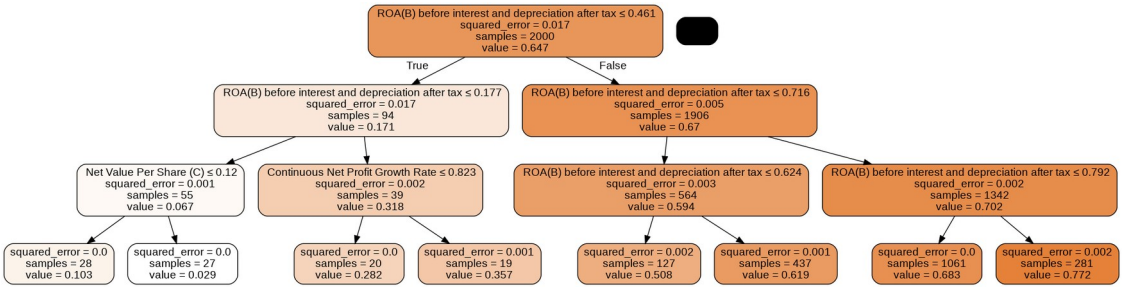
```
(0.004220688840176174, 3.171832005009734e-05)
```

Градиентный бустинг

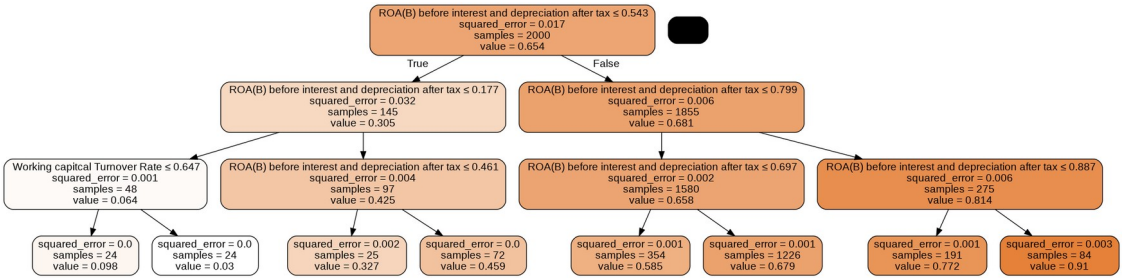
```
from sklearn.ensemble import AdaBoostRegressor
from sklearn.tree import DecisionTreeClassifier,
DecisionTreeRegressor, export_graphviz

ab1 = AdaBoostRegressor(n_estimators=4, random_state=2022)
ab1.fit(X, Y)

AdaBoostRegressor(n_estimators=4, random_state=2022)

# Визуализация дерева
def get_png_tree(tree_model_param, feature_names_param):
    dot_data = StringIO()
    export_graphviz(tree_model_param, out_file=dot_data,
feature_names=feature_names_param,
                    filled=True, rounded=True,
special_characters=True)
    graph = pydotplus.graph_from_dot_data(dot_data.getvalue())
    return graph.create_png()

Image(get_png_tree(ab1.estimators_[0], X.columns), width="500")
```



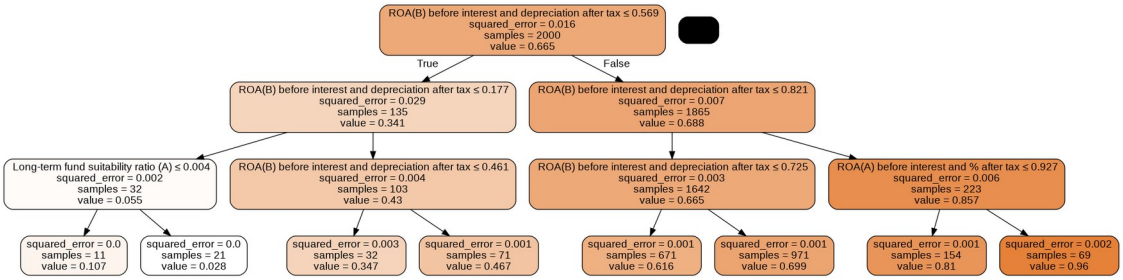

```
Image(get_png_tree(ab1.estimators_[1], X.columns), width="500")
```




```
Image(get_png_tree(ab1.estimators_[2], X.columns), width="500")
```




```
Image(get_png_tree(ab1.estimators_[3], X.columns), width="500")
```




```
regressor = AdaBoostRegressor(n_estimators=4, random_state=2022)
regressor.fit(X_train, Y_train)
y_pred = regressor.predict(X_test)
```

```
print('Mean Absolute Error:', mean_absolute_error(Y_test, y_pred))
print('Mean Squared Error:', mean_squared_error(Y_test, y_pred))
print('Root Mean Squared Error:', np.sqrt(mean_squared_error(Y_test,
y_pred)))
```

```
Mean Absolute Error: 0.012428410268481134
Mean Squared Error: 0.0002818729129094282
Root Mean Squared Error: 0.016789071234271067
```

Как видно, линейная регрессия показала более лучшие результаты, чем градиентный бустинг.