

# Модель прогнозирования кризиса гостиничного предприятия

В этом разделе мы построим две модели, прогнозирующие кризис гостиничного предприятия на основе финансовых коэффициентов, разработанных клиентом:

- 1) Коэффициент текущей ликвидности
- 2) Коэффициент срочной ликвидности
- 3) Коэффициент абсолютной ликвидности
- 4) Доля запасов в оборотных активах
- 5) Коэффициент автономии
- 6) Отношение заемных средств к собственным
- 7) DSCR
- 8) Маневренность капитала
- 9) Рентабельность продаж
- 10) Рентабельность активов
- 11) Рентабельность собственного капитала
- 12) Рентабельность инвестиций
- 13) Коэффициент фондоотдачи
- 14) Коэффициент оборачиваемости запасов
- 15) Коэффициент оборачиваемости дебиторской задолженности
- 16) Коэффициент оборачиваемости кредиторской задолженности

На основе этих финансовых коэффициентов мы попытаемся предсказать банкротство компании на следующий год. Результат предсказания такой модели может быть интерпретирован как мера кризисности финансовой ситуации в которой находится компания. Разумеется, не всякий кризис заканчивается ликвидацией предприятия, поэтому построенная положительный результат предсказания является индикатором тяжелого кризисного состояния, которое может привести к банкротству.

Помимо самих моделей, мы так же получим ранжирование финансовых показателей по степени влияния на оценку кризисности состояния предприятия. А как следствие — руководство к нормализации положения путем стабилизации наиболее важных финансовых коэффициентов.

Все построения и расчеты были проведены с использованием языка Python 3.8 и среды Jupyter-notebook, вместе с библиотеками numpy, pandas, sklearn, catboost.

## Данные

Для вычисления финансовых коэффициентов мы используем открытые данные РОССТАТ бухгалтерской отчетности за 4 года: с 2015 по 2018 г. Информацию о состоянии компаний — дата ликвидации, в случае если компания ликвидирована, мы возьмем из справочника ЕГРЮЛ с 2016 года по 2019 (нам нужны данные на год вперед от данных из РОССТАТ, так как мы прогнозируем банкротство только в следующем году).

Посмотрим на состав данных относительно прогнозируемой переменной: каков был процент компаний банкротов по годам.

Таблица 1:

	2015	2016	2017	2018
Количество компаний	4341	5270	5611	6243
Количество ликвидаций	133	131	137	234
Процент ликвидаций (в следующем году)	3%	2.5%	2.4%	3.7%

Мы видим увеличение общего числа предприятий и уровень числа ликвидированных каждый год около 3 %.

Посмотрим также на распределение некоторых финансовых коэффициентов в разрезе ликвидированных и не ликвидированных предприятий

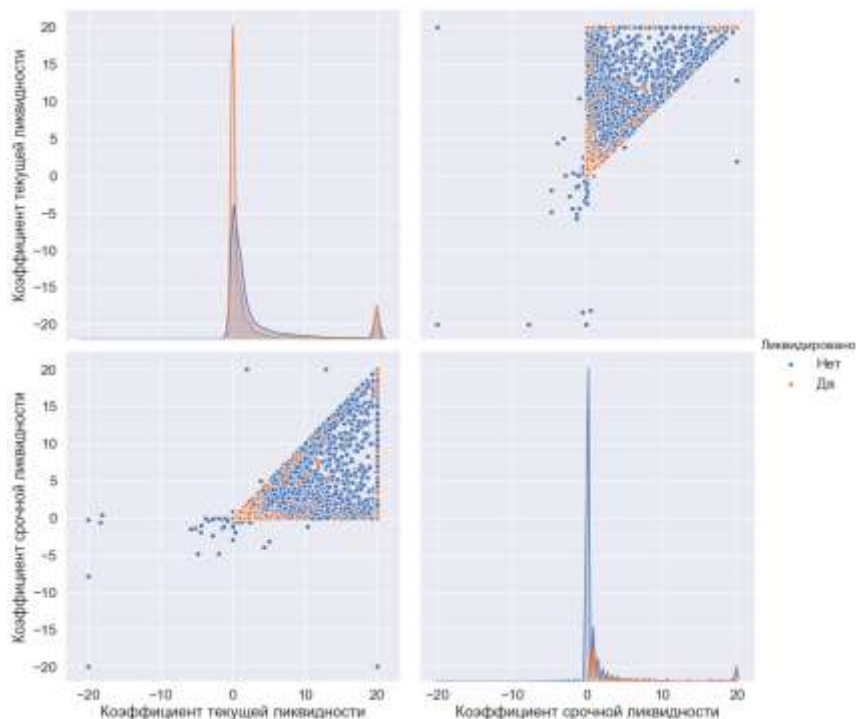


Рисунок 1 – Распределения коэффициентов текущей и срочной ликвидности для ликвидированных и не ликвидированных предприятий.

На графике изображено распределение коэффициентов текущей и срочной ликвидности для ликвидированных и не ликвидированных предприятий. В левом нижнем углу и правом верхнем каждой точке соответствует компании в координатах коэффициентов ликвидности. Оранжевые точки соответствуют ликвидированным компаниям. В левом верхнем и правом нижнем углу — распределение коэффициентов ликвидности. Легко заметить, что компаниям-банкротам (в будущем) соответствуют более низкие значения коэффициента текущей ликвидности и более высокие срочной ликвидности.

Аналогично ниже приведены распределения для пар (“Маневренность капитала”, “Рентабельность активов”).

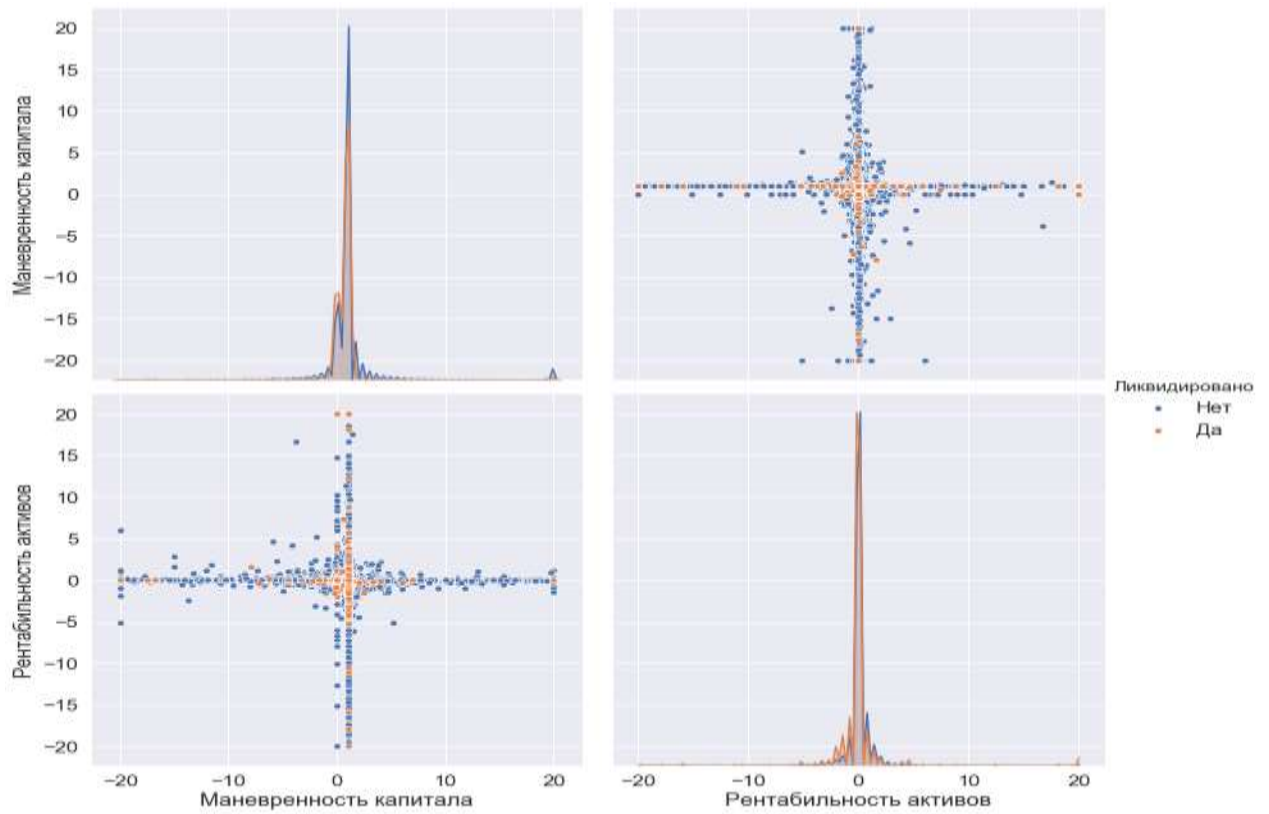


Рисунок 2 – Распределения для пар “Маневренность капитала”, “Рентабельность активов”.

А также (“DSCR коэффициент”, “Отношение собственных средств к заемным”):

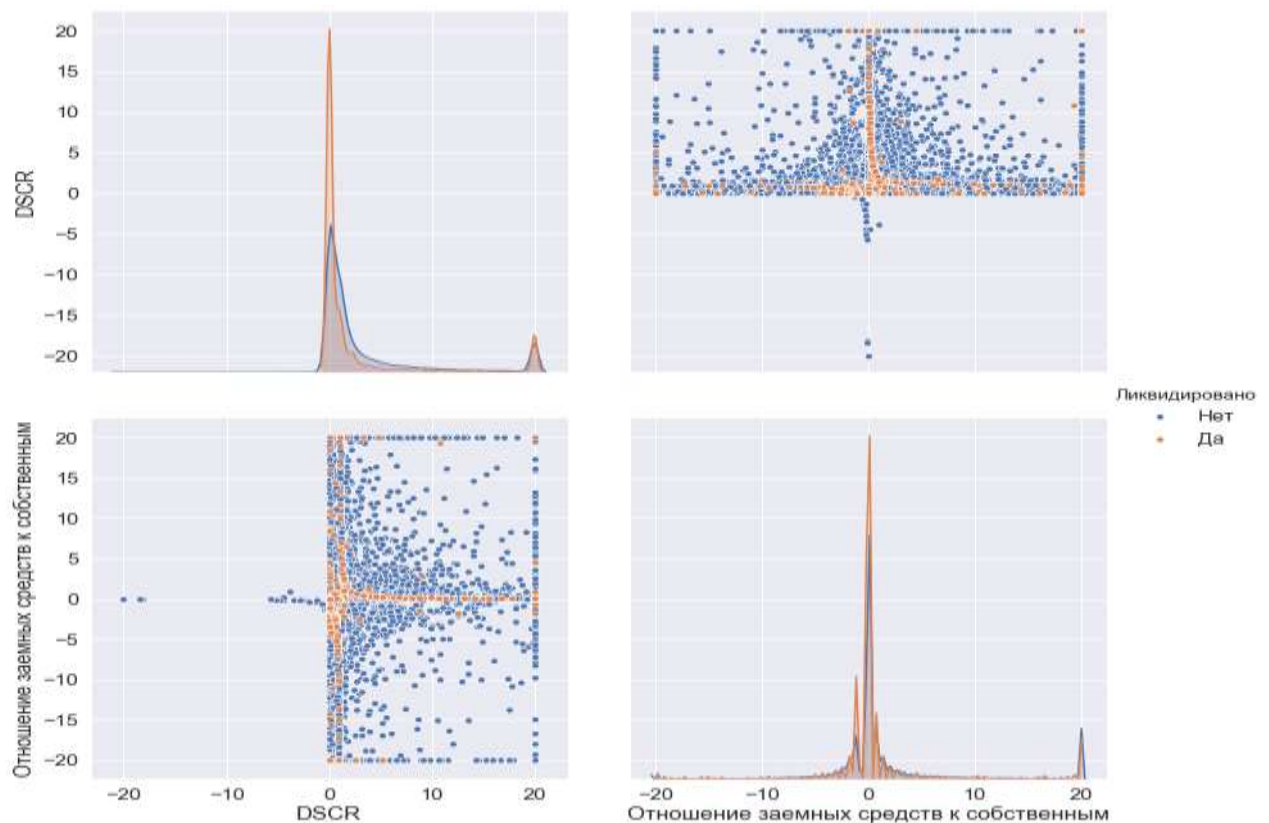


Рисунок 3 – Распределения для пар “DSCR коэффициент”, “Отношение собственных средств к заемным”.

Для нестабильных предприятий DSCR коэффициент так же ближе к 0 чем у стабильных. Рассмотрим корреляцию финансовых показателей друг с другом.

Как и ожидалось, сильно коррелированы между собой коэффициенты ликвидности, а также метрики, связанные с рентабельностью (за исключением рентабельности продаж). Отрицательно коррелированы маневренность капитала и отношение заемных средств к собственным. Также наблюдается строгая положительная корреляция между DSCR коэффициентом и показателями ликвидности. Положительно слабо коррелированы коэффициент автономии и рентабельность активов.

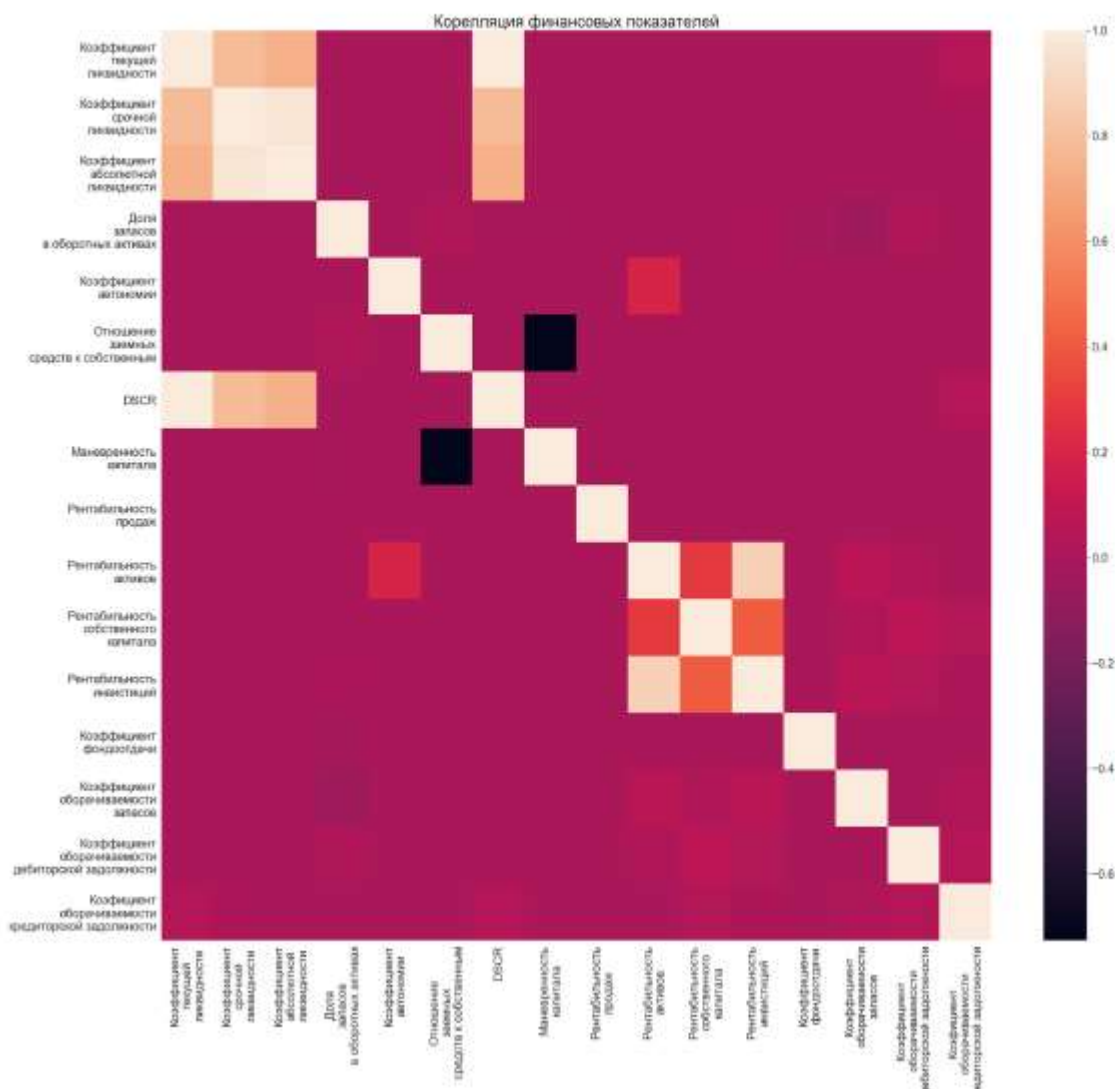


Рисунок 4 – Корреляция заданных финансовых показателей

# Модели

Для задачи аппроксимации данных, предсказания целевой переменной (факта ликвидации в нашем случае) существует большое множество математических методов, вот основные:

- 1) Линейные модели:
  - Линейная регрессия
  - Логистическая регрессия
  - “Лассо” регрессия
  - Метод опорных векторов
- 2) Метрические модели:
  - Метод ближайших соседей
- 3) Модели, основанные на деревьях решений:
  - Деревья решений
  - Случайный лес
  - Градиентный бустинг на основе деревьев решений
- 4) Нейронные сети:
  - Полносвязные нейронные сети
  - Сверточные нейронные сети
  - Рекуррентные нейронные сети

Это не весь перечень моделей и методов, однако список основных и наиболее распространенных. Преимуществом линейных моделей является простота и интерпретируемость, однако часто они не способны выявить глубокие нелинейные закономерности в данных. Интерпретируемой нелинейной моделью является дерево решений, однако оно обычно показывает низкое качество предсказаний. Более продвинутые методы на основе деревьев — случайный лес и градиентный бустинг. Особой популярностью пользуется последний. Так же есть нейронные сети, однако их применение целесообразно для данных чьи объемы и количество параметров в сотни раз превышают наши. К тому же, нейронные сети не интерпретируемые, другими словами — по построенной модели нельзя сказать, как каждый из показателей влияет на итоговый результат модели. Мы построили две модели.

## 1. Градиентный бустинг

Идея бустинг-подхода заключается в комбинации слабых (с невысокой обобщающей способностью) функций, которые строятся в ходе итеративного

процесса, где на каждом шаге новая модель обучается с использованием данных об ошибках предыдущих. Результирующая функция представляет собой линейную комбинацию базовых, слабых моделей. Будем рассматриваться бустинг деревьев решений: строить несколько деревьев, чтобы добавление новых деревьев уменьшало ошибку. При достаточно большом количестве деревьев мы сможем сильно уменьшить ошибку, однако не стоит забывать, что чем больше деревьев, тем дольше обучается модель и в какой-то момент прирост качества становится незначительным. В итоге наша модель будет иметь вид:

$$F(x) = \sum_{n=1}^N a_n D(x, p_n)$$

где  $a_n$  — коэффициенты,  $D$  — деревья решений, каждое из которых задается структурой  $p_n$ . Существует большое множество реализаций градиентного бустинга: XGBoost, lightgbm, CatBoos. Мы используем реализацию Catboost от Яндекса<sup>1</sup>. Ее преимуществом является автоматическая настройка гиперпараметров модели, при которых качество предсказания не уступает другим популярным реализациям, таким как XGBoost и lightgbm.

Мы не сможем выписать здесь саму модель в силу ее сложности (это комбинация порядка 300 деревьев решений). Однако можем привести результат и качество этой модели.

Задача, которую мы решаем — задача классификации, мы должны отнести каждое предприятие к одной из двух категорий, будет ли оно ликвидировано в следующем году или нет. Для измерения качества моделей в задачах классификации есть ряд метрик. Мы используем AUC-ROC — площадь под ROC-кривой.

ROC-кривая — график, позволяющий оценить качество бинарной классификации, отображает соотношение между долей объектов от общего количества носителей признака, верно классифицированных как несущие признак, и долей объектов от общего количества объектов, не несущих признака, ошибочно классифицированных как несущие признак при варьировании порога решающего правила.

Количественную интерпретацию ROC дает показатель AUC — площадь, ограниченная ROC-кривой и осью доли ложных положительных классификаций. Чем выше показатель AUC, тем качественнее классификатор, при этом значение 0,5 демонстрирует непригодность выбранного метода классификации (соответствует случайному гаданию). Значение менее 0,5 говорит, что классификатор действует с точностью до наоборот: если положительные назвать отрицательными и наоборот, классификатор будет работать лучше.

Простыми словами, данная величина есть вероятность того, что, выбрав наугад пару предприятий, мы верно скажем у какого предприятия выше вероятность банкротства. Мы выбрали эту метрику, а не более простую и общепринятую, как

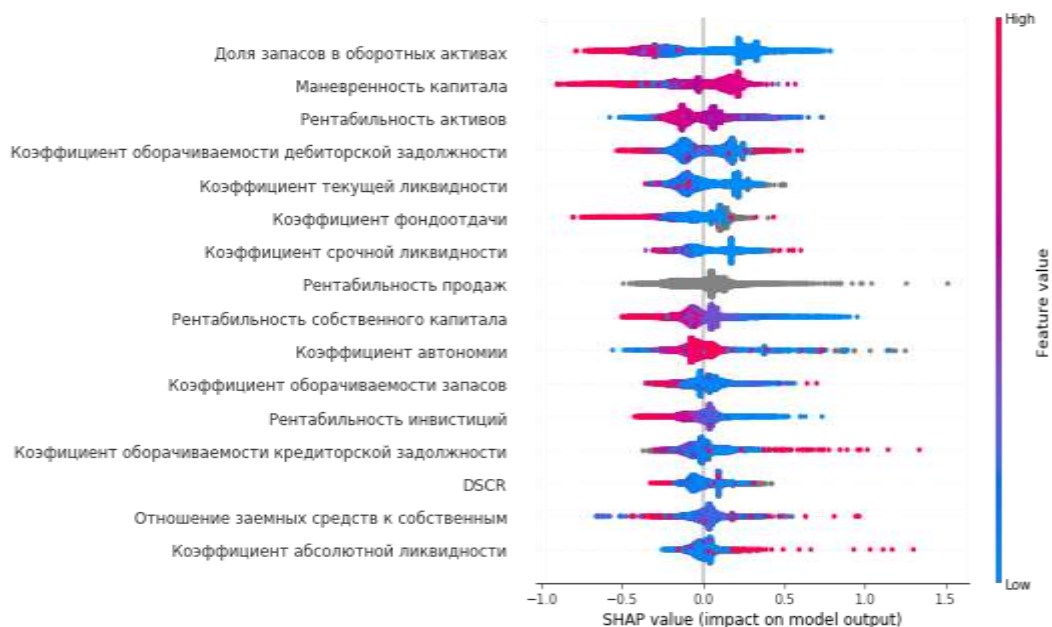
---

<sup>1</sup> <https://yandex.ru/dev/catboost/>.

например ассурасу (доля верных ответов модели), из-за несбалансированности целевой переменной: лишь 3 % гостиниц в итоге станут банкротами. Это означает, что если взять модель, которая выдает константный ответ — гостиница не обанкротится, то доля верных ответов будет 97 %, что разумеется не совпадает с реальным качеством модели. Метрика “Площадь под ROC кривой” лишена этого недостатка.

В результате обучения модели мы получили качество на тренировочной выборке: 0.86 и на тестовой выборке 0.75. Видно некоторое переобучение, однако модель на неизвестных ей данных дает 75% вероятности верного ранжирования гостиниц по их финансовому состоянию.

Помимо самой модели мы также можем получить значимость финансовых показателей при помощи SHAP-анализа. Это методика на основе теории игр, позволяющая дать оценку влияния параметров для любой модели-классификатора (см. например <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions>, [https://link.springer.com/chapter/10.1007/978-3-030-20055-8\\_4](https://link.springer.com/chapter/10.1007/978-3-030-20055-8_4).) Для построения SHAP значений мы использовали библиотеку <https://github.com/slundberg/shap>.



На данном графике финансовые показатели расположены сверху вниз по убыванию их влияния на вероятность банкротства. Каждая точка соответствует одному гостиничному предприятию. Цвет означает величину соответствующего признака: красный — данный признак выше среднего значения, синий — ниже. Отрицательное SHAP значение означает влияние данного признака в сторону уменьшения вероятности банкротства, положительное — в сторону увеличения.

Мы видим, что, как правило высокая доля запасов в оборотных активах, высокий коэффициент фондоотдачи, рентабельность собственного капитала —



положительно влияют на финансовое состояние. (Отрицательное SHAP значение в данном случае — означает уменьшение вероятности банкротства). Однако, достаточно часто наблюдается нелинейность: для одних предприятий высокое значение финансового параметра увеличивает вероятность банкротства, для других уменьшает. Особенно четко это видно на примере коэффициента оборачиваемости дебиторской задолженности (красные и синие точки есть по обе стороны от нуля), или маневренности капитала.

## 2. Логистическая регрессия.

Логистическая регрессия — это один из линейных методов прогнозирования. Это означает, что модель задается набором коэффициентов, и результирующий ответ — это линейная комбинация признаков объекта с соответствующими коэффициентами. Преимущество логистической регрессии в том, что при достаточно общих предположениях о распределении параметров объектов мы получаем точное значение вероятности принадлежности объекта одному из заданных классов: в нашем случае — это и есть вероятность того что компания будет ликвидирована в следующем году, либо не будет ликвидирована. Кроме того, логистическая регрессия, как и всякая линейная модель хорошо интерпретируема — мы можем сказать, как каждый из параметров влияет на вероятность банкротства. В итоге наша модель будет выглядеть следующим образом:

$$F(x) = f(a_0 + \sum_{n=1}^N a_n x_n)$$

Где  $a_i$  — коэффициенты модели,  $x_i$  — значения финансовых показателей гостиницы, а  $f(x) = \frac{1}{1 + e^{-x}}$  — преобразование, дающее на выход вероятность: число от 0 до 1.

В результате обучения модели мы получили следующую таблицу коэффициентов:

Таблица 2:

## Метрика

## Коэффициент

Коэффициент текущей ликвидности	-0,00171217838
Коэффициент срочной ликвидности	-0,0370059111
Коэффициент абсолютной ликвидности	0,009045252448
Доля запасов в оборотных активах	0,009045252448
Коэффициент автономии	-0,00171217838
Отношение заемных средств к собственным	-0,0370059111
DSCR	0,009045252448
Маневренность капитала	0,009045252448
Рентабельность продаж	-0,00171217838
Рентабельность активов	-0,0370059111
Рентабельность собственного капитала	0,009045252448
Рентабельность инвестиций	0,009045252448
Коэффициент фондоотдачи	-0,00171217838
Коэффициент оборачиваемости запасов	-0,0370059111
Коэффициент оборачиваемости дебиторской задолженности	0,009045252448
Коэффициент оборачиваемости кредиторской задолженности	0,009045252448
a_0 (Сдвиг)	-4,15

Знаки коэффициентов можно просто интерпретировать: отрицательное значение — увеличение данного финансового коэффициента негативно влияет на вероятность банкротства (уменьшает ее), положительный — наоборот. В целом знаки коэффициентов согласуются с интуитивными предположениями: высокие коэффициенты текущей и срочной ликвидности положительно влияют на общий финансовый профиль компании. Высокое же отношение заемных средств к собственным влияет отрицательно.

Измерим качество нашей модели тем же способом, что использовался для градиентного бустинга — площадь под ROC кривой. Значение качества на тренировочной выборке — 0.71, на тестовой — 0.69. Мы наблюдаем меньшее переобучение, чем в модели бустинга, но и качество ниже: 69% против 75% для бустинга. Это ожидаемый результат, так как логистическая регрессия не способна находить нелинейные закономерности, которые, как мы видели в SHAP анализе имеют место быть.

## Выводы

Мы увидели, что возможно спрогнозировать с 75% точностью кризис гостиничного предприятия на основании финансовых метрик. Нельзя судить о каждом показателе в отрыве от остальных в силу нелинейности найденной закономерности: как нам показал SHAP анализ, для одних предприятий высокое значение метрик увеличивает вероятность банкротства, а для других уменьшает. Это свидетельствует о необходимости комплексного анализа финансового состояния в контексте специфики данного гостиничного предприятия. Тем не менее любая из наших моделей может быть использована как единая комплексная метрика финансового благополучия гостиницы.

Помимо индикатора кризисного состояния, мы получили ранжирование финансовых метрик по степени важности. Разумеется, это лишь среднее значение для всей выборки гостиниц и в каждом отдельном случае необходим индивидуальный анализ. Однако при прочих равных, следует рассматривать финансовые показатели именно в том порядке, который нам дал SHAP анализ модели градиентного бустинга. Каждый из показателей влечет за собой список мер, направленных на то, чтоб приблизить его к среднему значению для гостиничных предприятий с благополучной финансовой ситуацией. Для заданного гостиничного предприятия, есть набор действий, которые тем или иным образом могут повлиять на финансовые метрики. Алгоритм применения модели будет следующий:

- 1) Для каждого из возможных действий оценить изменение финансовых метрик. Если прогнозирование метрики невозможно — оставить текущее значение.
- 2) Применить для каждого нового набора финансовых показателей модель градиентного бустинга (так как она показала лучшее качество чем логистическая регрессия).
- 3) Среди возможного множества действий, которые могут повлиять на финансовые показатели выбрать то, при котором предсказание модели наименьшее (наименьшая вероятность банкротства в следующем году).