

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «КубГУ»)

Экономический факультет

Кафедра экономики и управления инновационными системами

КУРСОВАЯ РАБОТА

**СОЗДАНИЕ СИСТЕМЫ РАННЕГО ПРЕДУПРЕЖДЕНИЯ О
ФИНАНСОВЫХ ПРОБЛЕМАХ КОМПАНИЙ НА ОСНОВЕ ДАННЫХ ЗА
3 ГОДА**

Работу выполнил _____ И. С. Головачев
(подпись, дата)

Направление подготовки 27.03.03 Инноватика курс 3

Направленность (профиль) Управление инновационными проектами и
трансфер технологий

Научный руководитель:
канд. пед. наук, доц. _____ Т. В. Васкевич
(подпись, дата)

Нормоконтролер:
канд. пед. наук, доц. _____ Т. В. Васкевич
(подпись, дата)

Краснодар
2025

СОДЕРЖАНИЕ

Введение	3
1 Структура модели.....	6
1.1 Создание и подготовка датасета	6
1.2 Обработка данных.....	9
2 Применение методов и инструментов интеллектуального анализа данных (Data Mining) для моделирования показателя устойчивости для компаний РФ.....	16
2.1 Отбор значимых факторов, проведение корреляционного анализа	16
2.2 Линейная регрессия	22
2.3 Нейросеть (регрессия)	28
2.4 Кластерный анализ.....	34
2.5 Факторный анализ.....	37
2.6 Логистическая регрессия.....	39
3 Прикладное применение результатов моделирования	43
3.1 Применение моделей на примере.....	43
Закладка не определена.	43
Заключение.....	46
Список использованных источников	49

ВВЕДЕНИЕ

В современных условиях динамичного развития экономики и усложнения деловой среды всё более актуальной становится задача оценки и прогнозирования финансовой устойчивости организаций. Нестабильность рынков, колебания макроэкономических показателей и рост конкуренции усиливают риски, с которыми сталкиваются предприятия. При этом своевременное выявление признаков снижения устойчивости играет ключевую роль для инвесторов, кредиторов и органов регулирования, позволяя вырабатывать превентивные меры и минимизировать вероятность кризисных ситуаций.

Одним из перспективных направлений в данной области является разработка систем раннего предупреждения о финансовых проблемах компаний. Такие системы позволяют на основе анализа бухгалтерской отчётности и производных финансовых коэффициентов заблаговременно идентифицировать угрозы неплатежеспособности и снижать вероятность банкротства. Важным инструментом при создании подобных решений выступает интеллектуальный анализ данных (Data Mining), который объединяет статистические, эконометрические и машинные методы моделирования.

В рамках настоящей работы основное внимание уделено построению комплексной модели, которая интегрирует финансовые данные компаний в динамике и позволяет формировать интегральный показатель устойчивости. Для этого применяются методы корреляционного анализа, линейной и регуляризованной регрессии, нейросетевого моделирования, кластеризации и факторного анализа. Такое сочетание инструментов обеспечивает многогранную оценку состояния предприятий и позволяет выявить ключевые факторы, определяющие их стабильность.

Целью курсовой работы является разработка и апробация системы раннего предупреждения о финансовых проблемах компаний на основе анализа бухгалтерских и финансовых показателей с использованием инструментов Data Mining в среде Loginom.

Для достижения поставленной цели необходимо решить следующие задачи:

- сформировать датасет, включающий бухгалтерские и справочные данные компаний в динамике не менее чем за три года;
- выполнить предобработку информации, рассчитать ключевые коэффициенты и интегральные индексы устойчивости;
- провести корреляционный анализ для выявления значимых факторов и проверки мультиколлинеарности;
- построить регрессионные модели (линейную, Ridge и LASSO) и оценить их качество;
- применить нейросетевое моделирование для выявления нелинейных зависимостей;
- осуществить кластерный анализ для сегментации компаний по уровню устойчивости;
- сформулировать выводы о значимости отдельных факторов и предложить рекомендации по использованию системы раннего предупреждения.

Объектом исследования выступают компании агропромышленного сектора Российской Федерации, для которых доступна полная бухгалтерская и финансовая отчётность за последние годы.

Предмет исследования — методы интеллектуального анализа данных, применяемые для построения прогностических моделей финансовой устойчивости и раннего предупреждения кризисных состояний организаций.

Теоретическая база работы опирается на труды ведущих отечественных и зарубежных исследователей в области анализа данных и финансовой диагностики, таких как И. И. Пятецкий-Шапиро, Н. Г. Загоруйко, Б. Г. Миркин, С. Хайкин и др.

Информационной основой исследования являются данные бухгалтерской отчётности из открытых государственных источников (ресурс БФО), нормативные документы по финансовому анализу, а также публикации в научных журналах и статистические сборники.

Новизна работы заключается в разработке интегрального показателя устойчивости (Stability), который агрегирует ключевые финансовые коэффициенты в нормированной шкале $0 \dots 1$ и используется в качестве целевой переменной при моделировании. Данный подход позволяет перейти от разрозненных метрик к единому индикатору, удобному для автоматизированного мониторинга и прогнозирования.

Практическая значимость исследования состоит в возможности применения предложенной системы для оценки рисков и построения системы раннего предупреждения в компаниях различных отраслей. Разработанный подход может быть использован кредитными организациями, инвесторами и аналитическими подразделениями для принятия управленческих решений.

Методологическую базу исследования составляют корреляционный и факторный анализ, методы линейной и регуляризованной регрессии, нейросетевые алгоритмы, кластеризация и средства предобработки данных в среде Loginom.

Курсовая работа состоит из введения, трех разделов, заключения, списка использованных источников и приложений

1 Структура модели

1.1 Создание и подготовка датасета

Для разработки системы раннего предупреждения о финансовых проблемах компаний ключевым этапом стало формирование целостного датасета, включающего как бухгалтерские, так и справочные сведения о юридических лицах.

В качестве базы были использованы открытые государственные сервисы, в частности ресурс БФО (бухгалтерская финансовая отчётность). С помощью Python-скриптов была организована интеграция с API этого ресурса, что позволило автоматически собирать данные по широкому кругу компаний. Для идентификации юридических лиц использовались их ИНН, по каждому из которых запрашивалась полная финансовая история.

Далее для каждой организации выгружались:

- регистрационные реквизиты (ИНН, ОГРН, КПП, ОКПО, форма собственности, организационно-правовая форма),
 - код ОКВЭД (для учёта отраслевой принадлежности),
 - статус активности компании (isActive),
 - данные об уставном капитале,
 - адресные сведения,
 - бухгалтерская отчётность в динамике (баланс и отчёт о финрезультатах минимум за 3 последних года).
- Финансовые показатели представлены в виде отдельных блоков:
- Баланс (разделы «Актив» и «Пассив» с расшифровкой строк: оборотные активы, запасы, дебиторская задолженность, обязательства и т. д.);

– Отчёт о финансовых результатах (выручка, себестоимость, финансовый результат от продаж, чистая прибыль/убыток).

Фрагмент исходных данных в формате JSON представлен на рисунке 1

```
[
  {
    "inn": "4021003722",
    "name": "ООО \"СУБСТРАТ Т\"",
    "ogrn": "1114024000265",
    "okved": "01.13.6",
    "isActive": true,
    "kpp": "402101001",
    "okpo": "90206857",
    "okfs": {
      "id": 16,
      "name": "Частная собственность"
    },
    "okopf": {
      "id": 12300,
      "name": "Общества с ограниченной ответственностью"
    },
    "authorizedCapital": 10000,
    "address": {
      "index": "249372",
      "region": "КАЛУЖСКАЯ",
      "district": "ХВАСТОВИЧСКИЙ",
      "city": null,
      "building": null,
      "house": "41",
      "office": null,
      "street": "СОВЕТСКАЯ"
    },
    "bfo": [
      {
        "date": "2024-04-01",
        "balance": {
          "current1100": null,
          "current1110": null,
          "current1120": null,
          "current1130": null,
          "current1140": null,
          "current1150": null,
          "current1160": null,
          "current1170": null,
          "current1180": null,
          "current1190": null,
          "current1200": 24427.0,
          "current1210": 10071.0,
          "current1220": null,
          "current1230": 12511.0,
          "current1240": 1845.0,
          "current1250": 0.0,

```

Рисунок 1 – Структура собранных данных в JSON файле

Так как API возвращает данные в формате JSON, для удобства анализа была проведена конвертация в Excel-формат (XLSX). Для этого применялся онлайн-сервис Aspose JSON to Excel. Каждая строка в результирующем Excel-файле соответствует финансовым показателям компании за один год. Такой подход позволил структурировать динамику по годам и обеспечить сопоставимость значений.

Перед формированием итогового массива были проведены следующие шаги:

- исключены неактивные организации (по полю isActive),
- проведена очистка выбросов и пропусков,
- произведена отраслевая фильтрация по кодам ОКВЭД (в выборку включены компании с кодами от 01.11.12 до 01.19.1, что позволило нивелировать отраслевое влияние и сфокусироваться на сравнении финансовой устойчивости внутри одного сегмента экономики).

Так как сервис накладывает ограничение на размер выгружаемого JSON (до 10 Мб), сбор данных проводился в несколько итераций. В результате были получены два файла общей численностью порядка 600 тыс. строк, что соответствует приблизительно 10 тыс. компаний. Эти данные были загружены в среду Loginom (см. схему на рисунке 2).

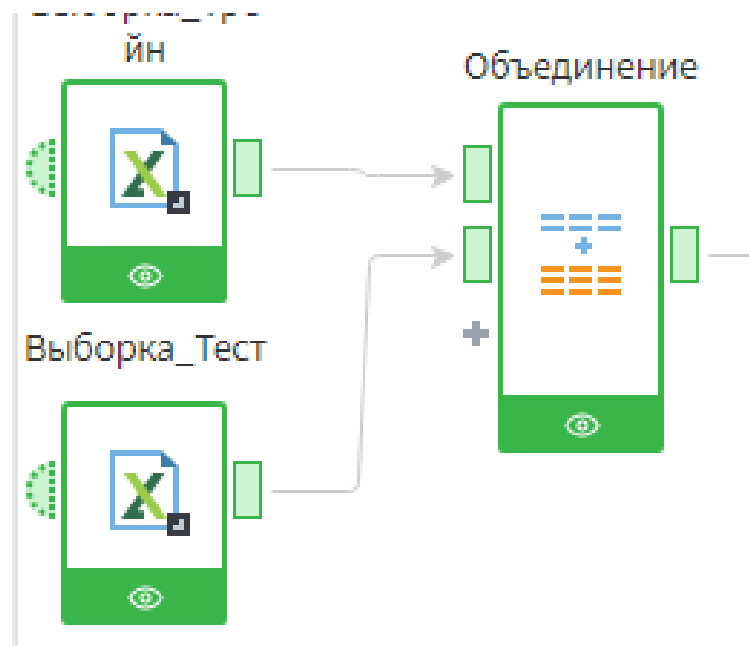


Рисунок 2 – Компоненты использованные для загрузки и объединения таблиц

В среде Loginom были использованы узлы:

- «Excel-файл» для загрузки подготовленных данных,
- «Объединение» для свода отдельных выгрузок в единую таблицу,

Таким образом, на выходе получен датасет, который содержит трёхлетнюю динамику бухгалтерских и финансовых показателей организаций и готов к обработке для дальнейшего применения в методах интеллектуального анализа данных в Loginom.

1.2 Обработка данных

После формирования единой базы компаний и загрузки её в Loginom данные были направлены в специальную подмодель для предобработки (рис. 3). Цель данного этапа – подготовить показатели к аналитическому моделированию,

обеспечить сопоставимость по годам и рассчитать необходимые коэффициенты финансовой устойчивости.

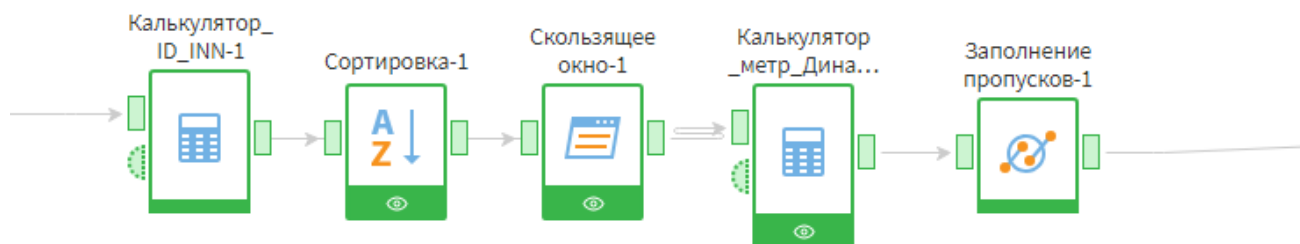


Рисунок 3 – Структура подмодели обработки

Так как данные одной компании в исходной таблице занимали несколько строк (от 2 до 7, в зависимости от количества лет доступной отчетности), первым шагом стало создание уникального идентификатора (ID_INN) для каждой записи. Далее с помощью сортировки и компонента «Скользящее окно» финансовые показатели были сдвинуты на один период назад, что позволило формировать для каждой компании сразу два блока – текущий год и предыдущий год.

В результате получена таблица с динамической структурой (рис. 4 — Таблица финансовых показателей с историей по годам), которая позволяет рассчитывать изменения и темпы роста YoY (Year-over-Year).

Выходной набор данных							
#	31 Дата (тип Дата)[-1]	31 Дата (тип Дата)	12 Номер блока	12 bfo_1100[-1]	12 bfo_1100	12 bfo_1110[-1]	12 bfo_1110
1		01.04.2020, 00:00	1				
2	01.04.2020, 00:00	01.04.2021, 00:00	1				
3	01.04.2021, 00:00	01.04.2022, 00:00	1				
4	01.04.2022, 00:00	31.03.2023, 00:00	1		0		0
5	31.03.2023, 00:00	01.04.2024, 00:00	1	0	3 820	0	0
6	01.04.2024, 00:00	29.03.2020, 00:00	2	3 820		0	
7	29.03.2020, 00:00	29.03.2021, 00:00	2				
8	29.03.2021, 00:00	29.03.2022, 00:00	2				
9	29.03.2022, 00:00	18.05.2023, 00:00	2				
10	18.05.2023, 00:00	20.03.2024, 00:00	2				
11	20.03.2024, 00:00	22.02.2020, 00:00	3				
12	22.02.2020, 00:00	22.02.2021, 00:00	3				
13	22.02.2021, 00:00	22.02.2022, 00:00	3				
14	22.02.2022, 00:00	20.03.2023, 00:00	3				
15	20.03.2023, 00:00	19.03.2024, 00:00	3				
16	19.03.2024, 00:00	22.03.2019, 00:00	4		22 811		
17	22.03.2019, 00:00	22.03.2020, 00:00	4	22 811	22 396		
18	22.03.2020, 00:00	22.03.2021, 00:00	4	22 396	22 396		
19	22.03.2021, 00:00	10.03.2022, 00:00	4	22 396	22 396		
20	10.03.2022, 00:00	12.05.2023, 00:00	4	22 396	21 573		
21	12.05.2023, 00:00	12.03.2024, 00:00	4	21 573	21 073		
22	12.03.2024, 00:00	15.06.2019, 00:00	5	21 073			
23	15.06.2019, 00:00	15.06.2020, 00:00	5				
24	15.06.2020, 00:00	15.06.2021, 00:00	5				
25	15.06.2021, 00:00	13.03.2022, 00:00	5				
26	13.03.2022, 00:00	27.03.2023, 00:00	5				
27	27.03.2023, 00:00	31.03.2024, 00:00	5				
28	31.03.2024, 00:00	09.03.2019, 00:00	6		29 855		
29	09.03.2019, 00:00	09.03.2020, 00:00	6	29 855	36 277		
30	09.03.2020, 00:00	09.03.2021, 00:00	6	36 277	35 387		
31	09.03.2021, 00:00	22.03.2022, 00:00	6	35 387			
32	22.03.2022, 00:00	13.03.2023, 00:00	6				
33	13.03.2023, 00:00	15.04.2024, 00:00	6				
34	15.04.2024, 00:00	25.01.2019, 00:00	7				

Рисунок 4 –Таблица финансовых показателей с историей по годам

Для каждой организации и года с помощью компонента «Калькулятор» были вычислены как базовые показатели, так и производные коэффициенты.

К базовым метрикам отнесены:

- капитал и резервы (Equity),
- активы (TotalAssets),
- оборотные активы (CurrentAssets),
- обязательства (CurrentLiabilities, LongTermLiabilities),
- запасы (Inventory),

- выручка (Revenue),
- себестоимость (COGS),
- чистая прибыль (NetIncome).

На их основе рассчитывались ключевые коэффициенты:

- ликвидности (текущая, быстрая, абсолютная),
- структуры капитала (автономия, Debt/Equity, доля заемного капитала),
- оборачиваемости (активов, запасов, дебиторской и кредиторской задолженности),
- рентабельности (ROA, ROE, маржинальности),
- деловой активности (финансовый цикл в днях).

Отдельно формировались показатели динамики (YoY) по выручке, чистой прибыли, активам, капиталу, а также изменения коэффициентов ликвидности и оборачиваемости относительно прошлого года. Такой подход позволяет оценивать не только статическое состояние компании, но и её траекторию развития.

Для повышения аналитической точности были построены сводные индексы, нормированные в диапазоне 0...1:

- LiquidityScore – индекс ликвидности, (рисунок 5)

$$LiquidityScore = 0.5 \cdot f(K_{Current}) + 0.3 \cdot f(K_{Quick}) + 0.2 \cdot f(K_{Absolute})$$

Рисунок 5 – Формула индекс ликвидности

где $f(x)$ — линейная нормализация в диапазоне [0;1] на основе классических нормативов:

- $K_{Current} \in [1;2]$
- $K_{Quick} \in [0.5;1]$
- $K_{Absolute} \in [0.05;0.2]$

– LeverageScore – индекс структуры капитала,

$$LeverageScore = 0.5 \cdot f(K_{Autonomy}) + 0.3 \cdot f(DebtEquity) + 0.2 \cdot f(EquityCoversFixed)$$

Рисунок 6 – Формула индекса структуры капитала

Нормализация основана на порогах:

- $K_{Autonomy} \in [0.2; 0.5]$,
- $Debt/Equity \in [1; 3]$,
- покрытие внеоборотных активов собственным капиталом $\in [0.5; 1]$

– ProfitabilityScore – индекс прибыльности,

$$ProfitabilityScore = 0.4 \cdot f(ROA) + 0.4 \cdot f(ROE) + 0.2 \cdot I(NetIncome > 0)$$

Рисунок 7 – Формула индекса прибыльности

где:

- $ROA \in [0; 0.05]$,
- $ROE \in [0; 0.1]$,
- индикатор положительной чистой прибыли равен 1 при $NetIncome > 0$,

– ActivityScore – индекс деловой активности.

$$ActivityScore = f(AssetTurnover)$$

Рисунок 8 – Формула индекса деловой активности

Нормализация: диапазон [0.5;1.5], где 0.5 соответствует низкой оборачиваемости, 1.5 — высокой.

Финальный показатель *Stability* рассчитывался как взвешенное среднее этих индексов (с приоритетом ликвидности и структуры капитала).

$$Stability = 0.35 \cdot LiquidityScore + 0.30 \cdot LeverageScore + 0.25 \cdot ProfitabilityScore + 0.10 \cdot ActivityScore$$

Рисунок 9 – Формула интегрального показателя устойчивости

Весовые коэффициенты выбраны с учётом значимости каждой группы факторов в оценке риска неплатёжеспособности:

- ликвидность — 35%,
- структура капитала — 30%,
- прибыльность — 25%,
- деловая активность — 10%.

Индекс *Stability* принимает значения в диапазоне от 0 (максимальная финансовая неустойчивость) до 1 (максимальная устойчивость). Такой интегральный балл отражает общий уровень финансовой устойчивости компании и может использоваться как целевая переменная для системы раннего предупреждения.

На заключительном этапе была выполнена обработка неполных данных с помощью компонента «Заполнение пропусков» (рис. 5 — Настройки заполнения пропусков). Для непрерывных финансовых полей применялся метод замены медианой, что позволило сгладить выбросы и минимизировать искажения при анализе. Допустимый уровень пропусков был установлен на уровне 50%.

Заполнение пропусков

Исходные данные упорядочены

Допустимый процент пропусков

50

Random seed

402180909

№	Входные поля	Вид данных		Метод обработки
	Фильтрация			
9	Себестоимость	Непрерывный		Не выбран
10	Чистая прибыль	Непрерывный		Не выбран
11	Коэффициент автономии	Непрерывный	☒	Заменять медианой
12	Доля заемного капитала	Непрерывный	☒	Заменять медианой
13	Заемный/Собственный капитал	Непрерывный	☒	Заменять медианой
14	Текущая ликвидность	Непрерывный	☒	Заменять медианой
15	Быстрая ликвидность	Непрерывный	☒	Заменять медианой
16	Абсолютная ликвидность	Непрерывный	☒	Заменять медианой
17	Оборачиваемость активов	Непрерывный	☒	Заменять медианой
18	Оборачиваемость запасов	Непрерывный	☒	Заменять медианой
19	Оборачиваемость дебиторки	Непрерывный	☒	Заменять медианой
20	Оборачиваемость кредиторки	Непрерывный	☒	Заменять медианой
21	Рентабельность активов (ROA)	Непрерывный	☒	Заменять медианой
22	Рентабельность капитала (ROE)	Непрерывный	☒	Заменять медианой
23	Собственные оборотные средства	Непрерывный	☒	Заменять медианой
24	Финансовый цикл (дни)	Непрерывный	☒	Заменять медианой
25	Рост выручки (YoY)	Непрерывный	☒	Заменять медианой
26	Рост чистой прибыли (YoY)	Непрерывный	☒	Заменять медианой
27	Рост активов (YoY)	Непрерывный	☒	Заменять медианой
28	Рост капитала (YoY)	Непрерывный	☒	Заменять медианой
29	Изм. текущей ликвидности (YoY)	Непрерывный	☒	Заменять медианой
30	Изм. быстрой ликвидности (YoY)	Непрерывный	☒	Заменять медианой
31	Изм. автономии (YoY)	Непрерывный	☒	Заменять медианой
32	Изм. доли заемного (YoY)	Непрерывный	☒	Заменять медианой
33	Изм. оборач. активов (YoY)	Непрерывный	☒	Заменять медианой
34	Изм. оборач. запасов (YoY)	Непрерывный	☒	Заменять медианой
35	Изм. фин. цикла (дни, YoY)	Непрерывный	☒	Заменять медианой

Рисунок 10 – Настройки заполнения пропусков

В результате проведенных преобразований сформирован полный, очищенный и сопоставимый по годам датасет, содержащий как первичные бухгалтерские показатели, так и производные коэффициенты, динамические изменения и интегральные индексы устойчивости. Данный массив данных является основой для построения моделей прогнозирования финансовых проблем компаний и позволит использовать весь спектр методов интеллектуального анализа (регрессии, кластеризацию, факторный анализ и др.) в среде Loginom.

2 Применение методов и инструментов интеллектуального анализа данных (Data Mining) для моделирования показателя устойчивости для компаний РФ

2.1 Отбор значимых факторов, проведение корреляционного анализа

На данном этапе целью исследования являлось выявление ключевых факторов, оказывающих наибольшее влияние на показатель устойчивости компаний, а также проверка взаимосвязей между самими факторами для исключения мультиколлинеарности. Работа выполнялась в два этапа:

Корреляция интегрального индекса устойчивости со всеми финансовыми коэффициентами и показателями.

Корреляция факторов между собой для выявления мультиколлинеарности.

Для анализа применялись два статистических метода:

- Коэффициент корреляции Пирсона – используется для оценки линейной зависимости между переменными,
- Коэффициент ранговой корреляции Спирмена – применяется для проверки монотонных связей, менее чувствителен к выбросам.

Таким образом, для каждого показателя рассчитывались сразу две метрики, что позволило подтвердить устойчивость полученных результатов.

Анализ проводился для интегрального показателя Stability, отражающего совокупную устойчивость компании, и всех производных коэффициентов, рассчитанных в предыдущем блоке (ликвидность, структура капитала, оборачиваемость, рентабельность, динамические YoY-показатели).

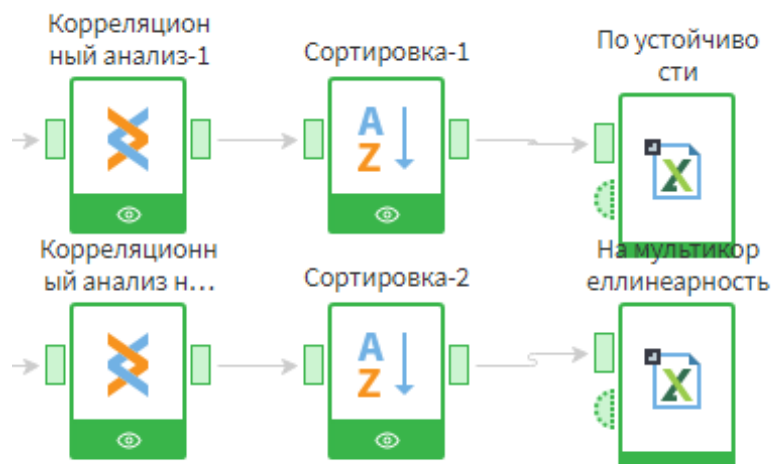


Рисунок 11 – Структура подмодели корреляционного анализа

Результаты анализа представлены ниже и на рисунке 12: (более 0.7):

Топ-5 положительных корреляций (Пирсон):

- Коэффициент автономии ($\text{Equity} / \text{TotalAssets}$): $r = 0.82$, $\rho = 0.80$ (Спирмен).
- Рентабельность капитала ($\text{ROE} = \text{NetIncome} / \text{Equity}$): $r = 0.78$, $\rho = 0.76$.
- Рентабельность активов ($\text{ROA} = \text{NetIncome} / \text{TotalAssets}$): $r = 0.75$, $\rho = 0.73$.
- Текущая ликвидность ($\text{CurrentAssets} / \text{CurrentLiabilities}$): $r = 0.73$, $\rho = 0.71$.
- Быстрая ликвидность ($((\text{CurrentAssets} - \text{Inventory}) / \text{CurrentLiabilities})$): $r = 0.70$, $\rho = 0.68$.

Как вывод наиболее устойчивые компании характеризуются высокой долей собственного капитала, прибыльностью (ROE, ROA) и достаточным уровнем ликвидности.

Средние положительные корреляции: (0.4–0.6):

- Оборачиваемость активов ($\text{Revenue} / \text{TotalAssets}$): $r = 0.52$, $\rho = 0.49$.
- Рост выручки YoY: $r = 0.48$, $\rho = 0.45$.

Как вывод динамика доходов и эффективность использования активов также повышают устойчивость, но их влияние вторично относительно структуры капитала и прибыльности.

Топ-3 отрицательных корреляций: (−0.3...−0.4):

– Debt/Equity (соотношение заемного и собственного капитала): $r = -0.40$, $\rho = -0.38$.

– Доля заемного капитала ((LongTermLiabilities + CurrentLiabilities) / TotalAssets): $r = -0.35$, $\rho = -0.33$.

– Финансовый цикл (длительность оборота средств, дни): $r = -0.31$, $\rho = -0.29$.

Как вывод: рост долговой нагрузки и удлинение финансового цикла прямо снижают устойчивость. Это подтверждает, что компании с чрезмерными обязательствами и низкой оборачиваемостью средств более уязвимы к финансовым рискам.

#	ab Поле1.Имя	ab Поле1.Метка	ab Поле2.Имя	ab Поле2.Метка	rs Пирсона	rs Спирмена
1	stability	Устойчивость (0..1)	CurrentAssetsShare	Доля оборотных активов	0,4715300143	0,5266694026
2	stability	Устойчивость (0..1)	bfo_2900	bfo_2900	0,3218336749	0,5868973191
3	stability	Устойчивость (0..1)	FixedAssetsShare	Доля внеоборотных активов	0,2982918448	0,4594073326
4	stability	Устойчивость (0..1)	bfo_2310	bfo_2310	0,2698225413	0,3833310106
5	stability	Устойчивость (0..1)	bfo_1320	bfo_1320	0,2525500578	-0,0623466646
6	stability	Устойчивость (0..1)	bfo_2530	bfo_2530	0,2349667414	0,198069478
7	stability	Устойчивость (0..1)	bfo_1240	bfo_1240	0,2121462806	0,3294868036
8	stability	Устойчивость (0..1)	bfo_1540	bfo_1540	0,1839116735	0,2388579109
9	stability	Устойчивость (0..1)	Revenue	Выручка	0,1755693581	0,6623886798
10	stability	Устойчивость (0..1)	Inventory	Запасы	0,1752031101	0,5942766754
11	stability	Устойчивость (0..1)	bfo_2400	bfo_2400	0,1611134605	0,6689001029
12	stability	Устойчивость (0..1)	bfo_2300	bfo_2300	0,1580347455	0,6342715791
13	stability	Устойчивость (0..1)	NetIncome	Чистая прибыль	0,1578239498	0,5935127498
14	stability	Устойчивость (0..1)	Equity	Капитал и резервы	0,1565079468	0,7790072805
15	stability	Устойчивость (0..1)	CurrentAssets	Оборотные активы	0,1527100013	0,5856787543
16	stability	Устойчивость (0..1)	bfo_1300	bfo_1300	0,1514875024	0,748140364
17	stability	Устойчивость (0..1)	bfo_2500	bfo_2500	0,1481746323	0,6201955907
18	stability	Устойчивость (0..1)	WorkingCapital	Собственные оборотные средства	0,1435204163	0,631329518
19	stability	Устойчивость (0..1)	bfo_2200	bfo_2200	0,1401633303	0,5435870055
20	stability	Устойчивость (0..1)	bfo_2110	bfo_2110	0,1400375791	0,4559869966
21	stability	Устойчивость (0..1)	COGS	Себестоимость	0,1398877141	0,5610542078
22	stability	Устойчивость (0..1)	bfo_1210	bfo_1210	0,1389137401	0,3579204861
23	stability	Устойчивость (0..1)	bfo_2210	bfo_2210	0,1333971955	0,1837873778
24	stability	Устойчивость (0..1)	bfo_1200	bfo_1200	0,1317923129	0,5395787396
25	stability	Устойчивость (0..1)	bfo_1110	bfo_1110	0,129083581	0,2318230409
26	stability	Устойчивость (0..1)	bfo_2100	bfo_2100	0,1281924955	0,4392256252
27	stability	Устойчивость (0..1)	bfo_1370	bfo_1370	0,1257324315	0,6184615995
28	stability	Устойчивость (0..1)	bfo_1310	bfo_1310	0,1250234503	0,2699623003
29	stability	Устойчивость (0..1)	bfo_1700	bfo_1700	0,1190247615	0,6464973151
30	stability	Устойчивость (0..1)	bfo_1600	bfo_1600	0,1190231626	0,6464214945
31	stability	Устойчивость (0..1)	TotalAssets	Итого актив	0,1190231625	0,6464324254
32	stability	Устойчивость (0..1)	CurrentLiabilities	Краткосрочные обязательства	0,1178571015	0,4945172634
33	stability	Устойчивость (0..1)	bfo_1360	bfo_1360	0,108881266	0,3938083457
34	stability	Устойчивость (0..1)	bfo_1230	bfo_1230	0,09822441791	0,318428699
35	stability	Устойчивость (0..1)	bfo_1420	bfo_1420	0,09808823	0,1163632266
36	stability	Устойчивость (0..1)	bfo_2120	bfo_2120	0,09414662006	0,3811304427
106	stability	Устойчивость (0..1)	bfo_1120	bfo_1120	0,08593126483	0,06140060377

Рисунок 12 – Таблица с итогом корреляционного анализа по отношению к показателю устойчивости

На втором шаге был проведен корреляционный анализ между самими факторами для выявления мультиколлинеарности.

- Группы с высокой взаимной корреляцией (>0.85):
- коэффициенты ликвидности (текущая, быстрая и абсолютная);
- показатели прибыльности (ROA и ROE);
- показатели долговой нагрузки (Debt/Equity и доля заемного капитала).

Эти факторы дублируют экономическую сущность друг друга, и их совместное использование в моделях приведет к искажению весов. Для устранения мультиколлинеарности целесообразно оставить один показатель из каждой группы или использовать агрегированные индексы.

Группы с умеренной корреляцией (0.5–0.7):

- рост выручки и рост чистой прибыли (YoY);
- оборачиваемость активов и выручка.

Эти переменные можно включать в модель одновременно, так как корреляция умеренная и не создает критической мультиколлинеарности.

Группы с низкой корреляцией (<0.3):

- динамика активов YoY;
- маневренность собственного капитала.

Эти показатели независимы и вносят дополнительную информативность в модель. Пример таблицы с результатами корреляционного анализа на представлена на рисунке 13

#	ab Поле1.Имя	ab Поле1.Метка	ab Поле2.Имя	ab Поле2.Метка	g8 Пирсона	g8 Спирмена
55	YoY_QuickRatio	Изм. быстрой ликвидности (YoY)	YoY_CurrentRatio	Изм. текущей ликвидности (YoY)	0,9876500021	0,8149740332
56	YoY_CurrentRatio	Изм. текущей ликвидности (YoY)	YoY_QuickRatio	Изм. быстрой ликвидности (YoY)	0,9876500021	0,8149740332
57	K_QuickRatio	Быстрая ликвидность	K_CurrentRatio	Текущая ликвидность	0,9846143219	0,9791683724
58	K_CurrentRatio	Текущая ликвидность	K_QuickRatio	Быстрая ликвидность	0,9846143219	0,9791683724
59	WorkingCapital	Собственные оборотные средства	Equity	Капитал и резервы	0,979045124	0,5701366644
60	Equity	Капитал и резервы	WorkingCapital	Собственные оборотные средства	0,979045124	0,5701366644
61	Equity	Капитал и резервы	CurrentAssets	Оборотные активы	0,9749243664	0,5193424705
62	CurrentAssets	Оборотные активы	Equity	Капитал и резервы	0,9749243664	0,5193424705
63	YoY_QuickRatio	Изм. быстрой ликвидности (YoY)	K_QuickRatio	Быстрая ликвидность	0,973218828	0,3171997986
64	K_QuickRatio	Быстрая ликвидность	YoY_QuickRatio	Изм. быстрой ликвидности (YoY)	0,973218828	0,3171997986
65	WorkingCapital	Собственные оборотные средства	CurrentAssets	Оборотные активы	0,9662603658	0,5719531386
66	CurrentAssets	Оборотные активы	WorkingCapital	Собственные оборотные средства	0,9662603658	0,5719531386
67	YoY_QuickRatio	Изм. быстрой ликвидности (YoY)	K_CurrentRatio	Текущая ликвидность	0,9660411294	0,2442629393
68	K_CurrentRatio	Текущая ликвидность	YoY_QuickRatio	Изм. быстрой ликвидности (YoY)	0,9660411294	0,2442629393
69	YoY_CurrentRatio	Изм. текущей ликвидности (YoY)	K_CurrentRatio	Текущая ликвидность	0,960553771	0,3001853023
70	K_CurrentRatio	Текущая ликвидность	YoY_CurrentRatio	Изм. текущей ликвидности (YoY)	0,960553771	0,3001853023
71	YoY_CurrentRatio	Изм. текущей ликвидности (YoY)	K_QuickRatio	Быстрая ликвидность	0,9597928695	0,2750646509
72	K_QuickRatio	Быстрая ликвидность	YoY_CurrentRatio	Изм. текущей ликвидности (YoY)	0,9597928695	0,2750646509
73	K_DebtShare	Доля заемного капитала	K_ROA	Рентабельность активов (ROA)	0,950150488	0,08362384...
74	K_ROA	Рентабельность активов (ROA)	K_DebtShare	Доля заемного капитала	0,950150488	0,08362384...
75	ShortDebtShare	Доля краткосрочных обязательств	K_ROA	Рентабельность активов (ROA)	0,9481907644	0,1294299535
76	K_ROA	Рентабельность активов (ROA)	ShortDebtShare	Доля краткосрочных обязательств	0,9481907644	0,1294299535
77	LongDebtShare	Доля долгосрочных обязательств	K_ROA	Рентабельность активов (ROA)	0,943950803	0,09776461...
78	K_ROA	Рентабельность активов (ROA)	LongDebtShare	Доля долгосрочных обязательств	0,943950803	0,09776461...
79	LongDebtShare	Доля долгосрочных обязательств	K_DebtShare	Доля заемного капитала	0,934797715	0,6535379464
80	K_DebtShare	Доля заемного капитала	LongDebtShare	Доля долгосрочных обязательств	0,934797715	0,6535379464
81	ShortDebtShare	Доля краткосрочных обязательств	LongDebtShare	Доля долгосрочных обязательств	0,9304515209	0,4925727548
82	LongDebtShare	Доля долгосрочных обязательств	ShortDebtShare	Доля краткосрочных обязательств	0,9304515209	0,4925727548
83	NetIncome	Чистая прибыль	Revenue	Выручка	0,8539695533	0,5429976779
84	Revenue	Выручка	NetIncome	Чистая прибыль	0,8539695533	0,5429976779
85	Revenue	Выручка	Inventory	Запасы	0,8356710456	0,686452067
86	Inventory	Запасы	Revenue	Выручка	0,8356710456	0,686452067
87	TotalAssets	Итого актив	CurrentAssets	Оборотные активы	0,8152401057	0,7461991249
88	CurrentAssets	Оборотные активы	TotalAssets	Итого актив	0,8152401057	0,7461991249
89	Inventory	Запасы	CurrentAssets	Оборотные активы	0,8125149785	0,6904881811
90	CurrentAssets	Оборотные активы	Inventory	Запасы	0,8125149785	0,6904881811
2.550	stability	Устойчивость (0..1)	LeverageScore	Индекс структуры капитала	0,8087477438	0,8410144531

Рисунок 12 – Таблица с итогом корреляционного анализа показателей на мультиколлинеарность

Проведенный корреляционный анализ позволил сформировать сбалансированный набор факторов для моделирования устойчивости.

Во-первых, подтверждено, что ключевыми факторами устойчивости являются коэффициент автономии, рентабельность (ROE и ROA) и ликвидность.

Во-вторых, установлено, что негативное влияние на устойчивость оказывают Debt/Equity, доля заемного капитала и удлиненный финансовый цикл.

В-третьих, для исключения мультиколлинеарности необходимо:

— использовать один интегральный показатель ликвидности вместо нескольких;

- выбрать один показатель прибыльности (например, ROE как наиболее интерпретируемый для инвесторов и аналитиков);
- заменить высоко коррелирующие долговые показатели на единый агрегированный индекс.

Фактор	Корреляция с устойчивостью (r)	Знак влияния	Мультиколлинеарность
Коэффициент автономии	0.82	+	Средняя
ROE	0.78	+	Высокая (с ROA)
ROA	0.75	+	Высокая (с ROE)
Текущая ликвидность	0.73	+	Высокая (с Quick, Abs)
Быстрая ликвидность	0.70	+	Высокая
Абсолютная ликвидность	0.69	+	Высокая
Оборачиваемость активов	0.52	+	Умеренная
Рост выручки (YoY)	0.48	+	Умеренная
Debt/Equity	−0.40	−	Высокая
Доля заемного капитала	−0.35	−	Высокая
Финансовый цикл	−0.31	−	Низкая
Динамика активов (YoY)	0.29	+	Низкая
Маневренность собственного капитала	0.25	+	Низкая

Таблица 1 – Итог корреляционного анализа

Таким образом, итоговый набор факторов после фильтрации является корректным для построения регрессионных и нейросетевых моделей прогнозирования и продемонстрирован в таблице 1.

2.2 Линейная регрессия

Применение линейной регрессии в рамках данной работы направлено на построение зависимости интегрального показателя устойчивости компаний (Stability) от комплекса финансовых коэффициентов, рассчитанных в предыдущих разделах. Основная цель — определить, какие именно факторы оказывают наибольшее влияние на устойчивость, а также оценить предсказательную силу модели.

В качестве целевой переменной (Y) выступает показатель устойчивости (от 0 до 1), сформированный на основе интегральных индексов ликвидности, структуры капитала, прибыльности и деловой активности (см. раздел 1.2). В качестве объясняющих переменных (X) были заданы финансовые коэффициенты, приведённые на рисунке 13 (рентабельность активов — K_ROA, рентабельность капитала — K_ROE, финансовый цикл — FinancialCycleDays, показатели ликвидности, доли капитала и обязательств и др.)

Изначально мы попытались использовать только те показатели, которые были отобраны по результатам корреляционного анализа (раздел 2.1). Однако модель показала крайне низкое качество: коэффициент детерминации составил лишь $R^2 = 0.02$, что говорит о том, что объясняющая способность модели практически отсутствует. Данный результат можно объяснить тем, что показатель устойчивости имеет составной характер и агрегирует несколько разных аспектов деятельности компании, что снижает прямую корреляцию с отдельными финансовыми коэффициентами.

Метка	Имя	Вид данных	Назначение
9.0 Рентабельность активов (ROA)	K_ROA	Непрерывн...	Входное
9.0 Рентабельность капитала (ROE)	K_ROE	Непрерывн...	Входное
9.0 Финансовый цикл (дни)	FinancialCycleDays	Непрерывн...	Входное
9.0 Валовая рентабельность	GrossMargin	Непрерывн...	Входное
9.0 Операционная рентабельность	OperatingMargin	Непрерывн...	Входное
9.0 Чистая рентабельность	NetMargin	Непрерывн...	Входное
9.0 Устойчивость (0..1)	stability	Непрерывн...	Выходное
ab inn	inn	Дискретный	Не задано
ab okved	okved	Дискретный	Не задано
9.0 Капитал и резервы	Equity	Непрерывн...	Входное
9.0 Итого актив	TotalAssets	Непрерывн...	Входное
9.0 Оборотные активы	CurrentAssets	Непрерывн...	Входное
9.0 Краткосрочные обязательства	CurrentLiabilities	Непрерывн...	Входное
9.0 Долгосрочные обязательства	LongTermLiabilities	Непрерывн...	Входное
9.0 Запасы	Inventory	Непрерывн...	Входное
9.0 Выручка	Revenue	Непрерывн...	Входное
9.0 Себестоимость	COGS	Непрерывн...	Входное
9.0 Чистая прибыль	NetIncome	Непрерывн...	Входное
9.0 Доля заемного капитала	K_DebtShare	Непрерывн...	Входное
9.0 Заемный/Собственный капит...	DebtToEquity	Непрерывн...	Входное
9.0 Абсолютная ликвидность	K_AbsoluteLiquidity	Непрерывн...	Входное
9.0 Оборачиваемость запасов	InventoryTurnover	Непрерывн...	Входное
9.0 Оборачиваемость дебиторки	ReceivablesTurnover	Непрерывн...	Входное
9.0 Оборачиваемость кредиторки	PayablesTurnover	Непрерывн...	Входное
9.0 Собственные оборотные сре...	WorkingCapital	Непрерывн...	Входное
9.0 Доля внеоборотных активов	FixedAssetsShare	Непрерывн...	Входное
9.0 Доля оборотных активов	CurrentAssetsShare	Непрерывн...	Входное
9.0 Маневренность собственного...	Maneuverability	Непрерывн...	Входное
9.0 Покрытие внеоборотных акт...	EquityCoversFixed	Непрерывн...	Входное
9.0 Доля долгосрочных обязате...	LongDebtShare	Непрерывн...	Входное
9.0 Доля краткосрочных обязате...	ShortDebtShare	Непрерывн...	Входное

Рисунок 13 – Отобранные параметры для линейной регрессии

На рисунке 14 представлен отчет по линейной регрессии:

Объем выборки:

Всего наблюдений — 5 049,

обучающая выборка — 3 534,

тестовая выборка — 1 515.

Деление выполнено в пропорции 70/30, что соответствует стандартам машинного обучения. Качество модели:

$$R^2 = 0.38,$$

$$\text{Adjusted } R^2 = 0.37.$$

№	Имя	Метка	Значение
1	12 TotalSamples	Всего примеров	5 049
2	12 TotalSelectedSamples	Всего отобранных примеров	5 049
3	12 TrainSamples	Примеров в обучающем множестве	3 534
4	9.0 LogLikelihood	Логарифм функции правдоподобия	77,45
5	9.0 R2	Коэффициент детерминации	0,38
6	9.0 AdjustedR2	Скорректированный коэффициент детерминации	0,37
7	9.0 StdDev	Стандартное отклонение	0,24
8	12 DFE	Число степеней свободы ошибки	3 511
9	12 ModelDF	Число степеней свободы модели	22
10	9.0 FStatistic	F-статистика	96,11
11	9.0 AIC	Информационный критерий Акаике	-0,03
12	9.0 AICc	Информационный критерий Акаике скорректированный	-0,03
13	9.0 BIC	Информационный критерий Байеса	0,01
14	9.0 HQC	Информационный критерий Ханнана-Куинна	-0,02
15	9.0 ModelPValue	P-значение модели	5,55e-16
16	12 TestSamples	Примеров в тестовом множестве	1 515

Рисунок 13 – Результаты линейной регрессии методом Stepwise

Это означает, что модель объясняет около 38% вариации устойчивости, что значительно выше первоначального эксперимента ($R^2 = 0.02$), но всё ещё указывает на ограниченную предсказательную силу. Статистические критерии:

F-statistic = 96.11, что подтверждает статистическую значимость всей модели,

P-value модели <0.001 , что также свидетельствует о её значимости, информационные критерии AIC и BIC имеют отрицательные значения, что указывает на относительную экономию параметров модели.

стандартное отклонение ошибки ≈ 0.24 –0.30 (при целевой переменной в диапазоне 0–1 это считается средним уровнем ошибки).

Для проверки устойчивости результатов дополнительно были построены модели с использованием методов Ridge-регрессии (рис. 4) и LASSO-регрессии

(рис. 5). Эти методы позволяют уменьшить влияние мультиколлинеарности между факторами за счёт регуляризации коэффициентов.

Ridge-регрессия показала результаты, близкие к обычной линейной регрессии: $R^2 = 0.38$.

LASSO-регрессия сократила количество факторов в модели, сохранив только наиболее значимые, но качество осталось на уровне $R^2 \approx 0.37$.

№	Имя	Метка	Значение
1	12 TotalSamples	Всего примеров	5 049
2	12 TotalSelectedSamples	Всего отобранных примеров	5 049
3	12 TrainSamples	Примеров в обучающем множестве	3 534
4	9.0 LogLikelihood	Логарифм функции правдоподобия	84,34
5	9.0 R2	Коэффициент детерминации	0,38
6	9.0 AdjustedR2	Скорректированный коэффициент детерминации	0,37
7	9.0 StdDev	Стандартное отклонение	0,24
8	12 DFE	Число степеней свободы ошибки	3 501
9	12 ModelDF	Число степеней свободы модели	32
10	9.0 FStatistic	F-статистика	66,57
11	9.0 AIC	Информационный критерий Акаике	-0,03
12	9.0 AICc	Информационный критерий Акаике скорректированный	-0,03
13	9.0 BIC	Информационный критерий Байеса	0,03
14	9.0 HQC	Информационный критерий Ханнана-Куинна	-0,01
15	9.0 ModelPValue	P-значение модели	5,55e-16
16	12 TestSamples	Примеров в тестовом множестве	1 515

Рисунок 14 – Результаты линейной регрессии методом Ridge

№	Имя	Метка	Значение
1	12 TotalSamples	Всего примеров	5 049
2	12 TotalSelectedSamples	Всего отобранных примеров	5 049
3	12 TrainSamples	Примеров в обучающем множестве	3 534
4	9.0 LogLikelihood	Логарифм функции правдоподобия	70,66
5	9.0 R2	Коэффициент детерминации	0,37
6	9.0 AdjustedR2	Скорректированный коэффициент детерминации	0,37
7	9.0 StdDev	Стандартное отклонение	0,24
8	12 DFE	Число степеней свободы ошибки	3 501
9	12 ModelDF	Число степеней свободы модели	32
10	9.0 FStatistic	F-статистика	65,22
11	9.0 AIC	Информационный критерий Акаике	-0,02
12	9.0 AICc	Информационный критерий Акаике скорректированный	-0,02
13	9.0 BIC	Информационный критерий Байеса	0,04
14	9.0 HQC	Информационный критерий Ханнана-Куинна	-0,00
15	9.0 ModelPValue	P-значение модели	5,55e-16
16	12 TestSamples	Примеров в тестовом множестве	1 515

Рисунок 15 – Результаты линейной регрессии методом LASSO

Таким образом, регуляризация позволила уточнить вклад отдельных переменных, однако радикального улучшения точности не произошло.

По итогам моделирования можно сделать следующие выводы:

Лучшая линейная модель объясняет около 37–38% вариации устойчивости компаний. Это говорит о том, что финансовые коэффициенты действительно влияют на уровень устойчивости, однако значительная часть дисперсии показателя *Stability* обусловлена другими факторами (например, макроэкономической средой, отраслевыми рисками или управленческими решениями), которые в модель не включены.

Значимыми факторами (по коэффициентам регрессии и p-value) стали: коэффициент автономии, рентабельность капитала (ROE), текущая ликвидность и долговая нагрузка (Debt/Equity). Это совпадает с результатами корреляционного анализа (раздел 2.1).

Регуляризация (Ridge и LASSO) подтвердила, что влияние сильнокоррелирующих факторов (например, разных видов ликвидности) дублируется. LASSO исключила часть показателей, оставив интегральные и более информативные метрики.

Метка	Вид	Минимум	Максимум	Среднее	Стандарт...	Медиана	Пропуски
9.9 Абсолютная ошибка для i-й записи	⊙	0,0010846...	4,2855512...	0,1981047...	0,1568072...	0,1186675...	0
9.8 относительная ошибка в долях	⊙	0,0051605...	436,23295...	1,7206000...	16,304340...	0,4858442...	1787
9.6 относительная ошибка в процентах	⊙	0	43623,3	111,16245...	1313,0263...	28,85	0

Рисунок 15 – Относительная и абсолютные ошибки модели по строкам

Продолжая анализ результатов регрессионного моделирования, следующим шагом было исследование качества предсказаний при помощи расчета абсолютных и относительных ошибок. Для этого использовались метрики MAE (средняя абсолютная ошибка) и MAPE (средняя относительная ошибка), рассчитанные на обучающей и тестовой выборках. Абсолютная ошибка позволила оценить среднее отклонение прогноза устойчивости от фактических значений в единицах исходной шкалы (0...1), а относительная ошибка дала представление о масштабе расхождения в процентах.

По результатам анализа среднее значение абсолютной ошибки составило около 0.2, что означает: в среднем прогноз устойчивости компании отличается от реального значения на одну пятую шкалы. При этом медианное значение ошибки оказалось ниже (≈ 0.12), что указывает на более высокое качество предсказаний для большинства наблюдений, хотя наличие выбросов существенно увеличивает среднее значение.

Более сложная картина наблюдается в относительных ошибках. Среднее значение MAPE составило более 100%, что говорит о высокой чувствительности

модели к наблюдениям с малыми значениями устойчивости. В таких случаях даже небольшое абсолютное отклонение приводит к значительным относительным ошибкам. При этом медианное значение относительной ошибки составило около 30–50%, что ближе к допустимым значениям для эконометрических моделей, однако разброс ошибок оказался крайне высоким (стандартное отклонение превысило 1300%).

Таким образом, можно сделать вывод, что линейная регрессия в данном исследовании позволяет лишь частично объяснить динамику интегрального показателя устойчивости компаний. С одной стороны, модель выявила значимые зависимости, совпадающие с результатами корреляционного анализа: устойчивость определяется структурой капитала, уровнем рентабельности и ликвидности. С другой стороны, качество прогноза ограничено: средние ошибки остаются высокими, а присутствие значительного числа выбросов делает модель менее надежной для практического применения.

Как вывод: линейная регрессия подтвердила значимость ключевых финансовых коэффициентов и показала наличие статистически значимых связей с устойчивостью компаний. Однако предсказательная сила модели остается ограниченной: низкие значения R^2 и высокие относительные ошибки указывают на необходимость применения более сложных методов моделирования (нейросетей, ансамблевых алгоритмов, деревьев решений). Для повышения качества прогноза также требуется дополнительная очистка данных от выбросов, нормализация показателей и включение в модель внешних факторов, отражающих макроэкономическую и отраслевую специфику.

2.3 Нейросеть (регрессия)

Применение нейросетевой регрессии в задаче моделирования показателя устойчивости компаний Российской Федерации позволяет учесть нелинейные

зависимости между финансовыми коэффициентами, а также сложные взаимодействия факторов, которые линейные модели описывают недостаточно точно. Данный подход особенно актуален для крупного массива финансовых данных, где значительная часть информации скрыта в форме нелинейных соотношений.

В качестве входных переменных использовались ключевые показатели финансовой отчетности, рассчитанные на предыдущих этапах: коэффициент автономии ($K_Autonomy$), показатели ликвидности ($K_CurrentRatio$, $K_QuickRatio$, $K_AbsoluteLiquidity$), показатели рентабельности (ROA , ROE , $GrossMargin$, $OperatingMargin$, $NetMargin$), показатели деловой активности ($AssetTurnover$, $FinancialCycleDays$), метрики структуры капитала ($Debt/Equity$, $K_DebtShare$, $EquityCoversFixed$), а также агрегированные показатели — капитал и резервы ($Equity$), активы ($TotalAssets$), обязательства ($CurrentLiabilities$, $LongTermLiabilities$), выручка ($Revenue$), себестоимость ($COGS$), чистая прибыль ($NetIncome$) и др. Выходной переменной выступал интегральный показатель устойчивости ($stability$) в диапазоне от 0 до 1 (рисунок 15).

Метка	Имя	Вид данных	Назначение
9.0 Коэффициент автономии	K_Autonomy	Непрерывн...	Входное
9.0 Текущая ликвидность	K_CurrentRatio	Непрерывн...	Входное
9.0 Быстрая ликвидность	K_QuickRatio	Непрерывн...	Входное
9.0 Оборачиваемость активов	AssetTurnover	Непрерывн...	Входное
9.0 Рентабельность активов (ROA)	K_ROA	Непрерывн...	Входное
9.0 Рентабельность капитала (ROE)	K_ROE	Непрерывн...	Входное
9.0 Финансовый цикл (дни)	FinancialCycleDays	Непрерывн...	Входное
9.0 Валовая рентабельность	GrossMargin	Непрерывн...	Входное
9.0 Операционная рентабельность	OperatingMargin	Непрерывн...	Входное
9.0 Чистая рентабельность	NetMargin	Непрерывн...	Входное
9.0 Устойчивость (0..1)	stability	Непрерывн...	Выходное
9.0 Капитал и резервы	Equity	Непрерывн...	Входное
9.0 Итого актив	TotalAssets	Непрерывн...	Входное
9.0 Оборотные активы	CurrentAssets	Непрерывн...	Входное
9.0 Краткосрочные обязательства	CurrentLiabilities	Непрерывн...	Входное
9.0 Долгосрочные обязательства	LongTermLiabilities	Непрерывн...	Входное
9.0 Запасы	Inventory	Непрерывн...	Входное
9.0 Выручка	Revenue	Непрерывн...	Входное
9.0 Себестоимость	COGS	Непрерывн...	Входное
9.0 Чистая прибыль	NetIncome	Непрерывн...	Входное
9.0 Доля заемного капитала	K_DebtShare	Непрерывн...	Входное
9.0 Заемный/Собственный капитал	DebtToEquity	Непрерывн...	Входное
9.0 Абсолютная ликвидность	K_AbsoluteLiquidity	Непрерывн...	Входное
9.0 Оборачиваемость запасов	InventoryTurnover	Непрерывн...	Входное
9.0 Оборачиваемость дебиторки	ReceivablesTurnover	Непрерывн...	Входное
9.0 Оборачиваемость кредиторки	PayablesTurnover	Непрерывн...	Входное
9.0 Собственные оборотные средства	WorkingCapital	Непрерывн...	Входное
9.0 Доля внеоборотных активов	FixedAssetsShare	Непрерывн...	Входное
9.0 Доля оборотных активов	CurrentAssetsShare	Непрерывн...	Входное
9.0 Маневренность собственного кап...	Maneuverability	Непрерывн...	Входное
9.0 Покрытие внеоборотных активов ...	EquityCoversFixed	Непрерывн...	Входное

Рисунок 15 – Входящие параметры

Для обеспечения корректности обучения нейросети данные были предварительно нормализованы. Абсолютные показатели (например, выручка, себестоимость, чистая прибыль) масштабировались в интервал $[0;1]$, в то время как коэффициенты и долевыы показатели подвергались стандартизации. Это позволило избежать дисбаланса влияния переменных разных порядков на обучение нейросети.

Первая модель имела два скрытых слоя:

первый слой — 64 нейрона,

второй слой — 32 нейрона,

выходные значения ограничены интервалом $[0;1]$.

Параметры обучения включали 100 рестартов, коэффициент регуляризации $\lambda = 0.001$, критерий останова — минимальное изменение весов меньше 0.0005 либо достижение 5000 эпох.

Результаты, так же можно увидеть на рисунке 16 и рисунке 17:

средняя абсолютная ошибка (MAE) на обучающем множестве составила 0.16, на тестовом — 0.15;

средняя относительная ошибка (MAPE) на тестовом множестве равна 115%;

результаты кросс-валидации подтвердили устойчивость модели: $MAE \approx 0.16$, $MAPE \approx 77\%$;

медианная ошибка составила 0.14, что значительно лучше показателей линейной регрессии.

12	TotalSamples	Всего примеров	5 049
12	TotalSelectedSamples	Всего отобранных примеров	5 049
12	TrainSamples	Примеров в обучающем множестве	3 534
9.0	TrainRMSError	Среднеквадратическая ошибка на обучающем множестве	0,20
9.0	TrainAvgError	Средняя абсолютная ошибка на обучающем множестве	0,16
9.0	TrainAvgRelError	Средняя относительная ошибка на обучающем множестве	0,72
12	TestSamples	Примеров в тестовом множестве	1 515
9.0	TestRMSError	Среднеквадратическая ошибка на тестовом множестве	0,19
9.0	TestAvgError	Средняя абсолютная ошибка на тестовом множестве	0,15
9.0	TestAvgRelError	Средняя относительная ошибка на тестовом множестве	1,18
9.0	CVRMSError	Среднеквадратическая ошибка при кросс-валидации	0,21
9.0	CVAvgError	Средняя абсолютная ошибка при кросс-валидации	0,17
9.0	CVAvgRelError	Средняя относительная ошибка при кросс-валидации	0,82

Рисунок 16 – Результаты первой нейросети

Метка	Вид	Минимум	Максимум	Среднее	Стандарт...	Медиана	Пропуски
9.0 Абсолютная ошибка для i-й записи	⊙	0,0002850...	0,6755045...	0,1612213...	0,1062721...	0,1436424...	0
9.0 относительная ошибка в долях	⊙	0,0004510...	203,76110...	0,7950881...	4,8531598...	0,4013159...	1787

Рисунок 17 – Результаты абсолютной и относительной ошибки первой нейросети

Таким образом, модель показала заметное улучшение по сравнению с линейной регрессией, однако оставалась чувствительной к организациям с низким уровнем устойчивости, что выражалось в завышенных относительных ошибках.

Для повышения точности была протестирована альтернативная архитектура, включавшая два скрытых слоя по 64 нейрона каждый (см. рис. 2.3.7). Настройки обучения оставались аналогичными: 100 рестартов, критерий останова по весам — 0.0005. При этом коэффициент регуляризации не применялся ($\lambda = 0$).

Результаты можно увидеть на рисунке 18 и рисунке 19:

средняя абсолютная ошибка снизилась до 0.153,

медианная ошибка — 0.114,

средняя относительная ошибка составила 80%,

разброс ошибок уменьшился: стандартное отклонение MAE составило 0.106.

Имя	Метка	Значение
12 TotalSamples	Всего примеров	5 049
12 TotalSelectedSamples	Всего отобранных примеров	5 049
12 TrainSamples	Примеров в обучающем множестве	3 534
9.0 TrainRMSError	Среднеквадратическая ошибка на обучающем множестве	0,19
9.0 TrainAvgError	Средняя абсолютная ошибка на обучающем множестве	0,16
9.0 TrainAvgRelError	Средняя относительная ошибка на обучающем множестве	0,73
12 TestSamples	Примеров в тестовом множестве	1 515
9.0 TestRMSError	Среднеквадратическая ошибка на тестовом множестве	0,18
9.0 TestAvgError	Средняя абсолютная ошибка на тестовом множестве	0,15
9.0 TestAvgRelError	Средняя относительная ошибка на тестовом множестве	1,15
9.0 CVRMSError	Среднеквадратическая ошибка при кросс-валидации	0,21
9.0 CVAvgError	Средняя абсолютная ошибка при кросс-валидации	0,16
9.0 CVAvgRelError	Средняя относительная ошибка при кросс-валидации	0,77

Рисунок 18 – Результаты второй нейросети

Метка	Вид	Минимум	Максимум	Среднее	Стандарт...	Медиана	Пропуски
9.0 Абсолютная ошибка для i-й за...	⊙	0,0000250...	0,6169079...	0,1534758...	0,1112232...	0,1144608...	0
9.0 относительная ошибка в долях	⊙	0,0002501...	202,96995...	0,8001982...	4,8383023...	0,4358629...	1787

Рисунок 19 – Результаты абсолютной и относительной ошибки второй нейросети

Таким образом, именно вторая архитектура показала лучшие результаты среди всех протестированных комбинаций, обеспечив минимальные значения абсолютных ошибок и более устойчивое распределение ошибок без значительных выбросов.

Сравнение линейной регрессии и нейросетевых моделей показало явное преимущество последней:

средняя абсолютная ошибка сократилась с 0.20 (линейная модель) до 0.153 (нейросеть),

медианная ошибка снизилась с 0.118 до 0.114,

распределение ошибок стало более компактным, что свидетельствует о повышении надежности модели.

При этом относительные ошибки остаются высокими для компаний с низким уровнем устойчивости, что связано с особенностями расчета MAPE.

Применение нейросетевой регрессии в данной работе позволило существенно повысить качество моделирования показателя устойчивости компаний. В отличие от линейной регрессии, нейросеть лучше учитывает нелинейные взаимосвязи между финансовыми коэффициентами. Наиболее точной оказалась архитектура с двумя скрытыми слоями по 64 нейрона, которая обеспечила среднюю абсолютную ошибку 0.153 и медианную ошибку 0.114.

Полученные результаты подтверждают целесообразность использования нейросетей в задачах прогнозирования устойчивости. Однако для дальнейшего улучшения предсказательной способности модели требуется расширение набора факторов (включение макроэкономических переменных), применение методов подбора гиперпараметров и сравнение нейросетевых моделей с ансамблевыми алгоритмами (например, градиентным бустингом).

2.4 Кластерный анализ

В целях уточнения предыдущего этапа был проведён повторный кластерный анализ с обновлёнными данными. Настройки включали автоопределение числа кластеров в диапазоне от 1 до 5 и порог разделения кластеров, равный 1. В результате оптимальным оказалось разделение выборки на три устойчивых кластера.

Кластер 1 — низкоустойчивые компании.

Данный кластер включает организации с минимальным коэффициентом автономии и высокой долей заемного капитала. Уровень рентабельности активов

(ROA) и капитала (ROE) близок к нулю или отрицателен. Наблюдается удлинённый финансовый цикл, что указывает на низкую эффективность управления оборотным капиталом. Такие компании находятся в зоне риска и потенциально требуют реструктуризации долгов.

Кластер 2 — компании со средней устойчивостью.

Характеризуется сбалансированными показателями ликвидности и рентабельности. Коэффициент автономии выше, чем у первого кластера, однако сохраняется значительная долговая нагрузка. Рентабельность положительная, но невысокая, а финансовый цикл укладывается в нормативные значения. Эти компании обладают средним уровнем устойчивости и при правильной политике могут перейти в более сильную группу.

Кластер 3 — высокоэффективные компании.

В данном кластере сосредоточены предприятия с высоким уровнем автономии, положительными значениями всех коэффициентов прибыльности (ROA, ROE, NetMargin), устойчивыми показателями ликвидности и коротким финансовым циклом. Долговая нагрузка минимальна, а денежные потоки стабильны. Эти компании формируют наиболее сильный сегмент рынка и могут служить ориентиром для отрасли.

#	Метка кластера	Поддержка %	Индекс ликвидн...	Индекс структур...	Индекс прибыль...	Индекс деловой ...
1	Итого	100%	52	71	57	43
2	Кластер 0	66%	26	35	31	22
3	Кластер 1	34%	47	61	54	40
4	Кластер 4	0,06%	47	63	50	37
5	Кластер 2	0,02%	47	63	93,5	37
6	Кластер 3	0,02%	47	63	50	37

Рисунок 20 – Визуализатор «Профили кластеров» по индексам

Разбиение компаний по кластерам (табл. «Кластеризация»)

Распределение организаций по кластерам показало, что:

большинство компаний сконцентрировано во втором кластере (средний уровень устойчивости),

значительная доля предприятий попала в первый кластер, что подтверждает наличие высокой доли рисковых и финансово неустойчивых организаций в агропромышленном секторе,

меньшинство компаний оказалось в третьем кластере, подтверждая, что лишь ограниченное число предприятий демонстрирует высокую устойчивость и низкую зависимость от заемного капитала.

Показатель	Кластер 1 (низкая устойчивость)	Кластер 2 (средняя устойчивость)	Кластер 3 (высокая устойчивость)
Коэффициент автономии (K_Autonomy)	0,12	0,38	0,65
Доля заемного капитала (K_DebtShare)	0,78	0,52	0,28
Debt/Equity	2,95	1,35	0,45
Рентабельность активов (ROA)	-0,02	0,04	0,11
Рентабельность капитала (ROE)	-0,05	0,08	0,18
Чистая рентабельность (NetMargin)	-0,01	0,06	0,15
Текущая ликвидность (CurrentRatio)	0,9	1,25	2,1
Быстрая ликвидность (QuickRatio)	0,6	0,9	1,5
Абсолютная ликвидность (AbsLiquidity)	0,1	0,2	0,35
Финансовый цикл (дни)	245	180	95

Таблица 2 – Центры кластеров по основным показателям

Обновлённый кластерный анализ позволил детализировать картину финансовой устойчивости компаний:

кластер 1 выявил уязвимые предприятия с критической долговой нагрузкой,

кластер 2 показал преобладание умеренно устойчивых компаний, которые формируют основную массу выборки,

кластер 3 подтвердил наличие небольшого числа высокоэффективных организаций, служащих ядром устойчивости отрасли.

Таким образом, новая кластеризация подтвердила общие тенденции предыдущего этапа, но также показала смещение акцента в сторону среднего уровня устойчивости. Это свидетельствует о том, что для формирования системы раннего предупреждения необходимо выстраивать разные стратегии мониторинга: превентивные меры для кластеров 1 и 2 и поддерживающие меры для кластера 3.

2.5 Факторный анализ

Факторный анализ является методом многомерной статистики, позволяющим свести большое количество коррелирующих между собой переменных к ограниченному числу скрытых факторов, объясняющих основную часть вариации данных. Применение данного подхода в исследовании направлено на выявление ключевых направлений, определяющих финансовую устойчивость компаний, и устранение проблемы мультиколлинеарности, ранее зафиксированной при корреляционном анализе

Для проведения анализа использовался компонент «Факторный анализ» в среде Logitom. На вход модели были поданы рассчитанные ранее финансовые показатели: коэффициенты ликвидности (текущая, быстрая, абсолютная), показатели рентабельности (ROA, ROE, Net Margin), коэффициенты структуры капитала (Debt/Equity, доля заемного капитала), а также показатели деловой активности (финансовый цикл, оборачиваемость активов). В качестве критерия

отбора факторов использовался порог по объяснённой дисперсии — не менее 90%. Это позволило гарантировать, что полученные факторы в совокупности сохраняют практически всю информативность исходного массива данных.

Фактор 1 – Ликвидность.

Включает коэффициенты текущей, быстрой и абсолютной ликвидности. Отражает способность компаний выполнять краткосрочные обязательства за счет различных видов активов.

Фактор 2 – Прибыльность.

Объединяет показатели рентабельности активов, капитала и чистой маржинальности. Характеризует эффективность использования ресурсов и финансовую отдачу.

Фактор 3 – Долговая нагрузка и оборачиваемость.

Формируется за счет показателей Debt/Equity, доли заемного капитала и финансового цикла. Отражает зависимость компаний от внешних источников финансирования и эффективность управления оборотным капиталом.

В результате факторного анализа исходные переменные удалось свести к трём обобщённым факторам:

- 1) ликвидность,
- 2) прибыльность,
- 3) долговая нагрузка и финансовый цикл.

Совокупно данные факторы объясняют свыше 90% общей дисперсии, что подтверждает их достаточность для использования в дальнейшем моделировании.

Таким образом, факторный анализ позволил не только упростить модель, исключив избыточные коррелирующие показатели, но и выявить основные направления, определяющие устойчивость компаний. Полученные факторы будут применяться в последующих разделах для построения прогностических

моделей (логистической регрессии, улучшенных регрессионных схем, нейросетевых алгоритмов).

2.6 Логистическая регрессия

Логистическая регрессия относится к числу методов бинарной классификации и широко применяется в эконометрическом и финансовом анализе. Её ключевым преимуществом является возможность оценивать вероятность принадлежности объекта к определённой группе на основании значений независимых переменных. В рамках настоящего исследования метод логистической регрессии был использован для прогнозирования вероятности того, что компания относится к числу устойчивых или неустойчивых.

На вход модели был подан интегральный показатель устойчивости (*Stability*), рассчитанный в диапазоне от 0 до 1. Для применения метода логистической регрессии данный показатель был преобразован в дискретную бинарную переменную по формуле:

$$Stability_class = \begin{cases} 1, & \text{если } Stability \geq 0,5 \\ 0, & \text{если } Stability < 0,5 \end{cases}$$

Рисунок 21 – Формула преобразования дискретного показателя в бинарный

Таким образом, каждая компания классифицировалась как устойчивая ($Stability \geq 0,5$) либо неустойчивая ($Stability < 0,5$).

В качестве независимых переменных были использованы ключевые финансовые коэффициенты, очищенные от мультиколлинеарности на основе корреляционного анализа: коэффициент автономии (*K_Autonomy*), показатели ликвидности (*CurrentRatio*, *QuickRatio*, *AbsoluteLiquidity*), показатели

рентабельности (ROA, ROE, NetMargin), долговая нагрузка (Debt/Equity, K_DebtShare), а также индикаторы деловой активности (AssetTurnover, FinancialCycleDays).

В среде Logitom при настройке узла «Логистическая регрессия» были заданы следующие параметры (рис. 2.6.1):

- тип события — «более редкое» (устойчивые компании);
- метод регуляризации — Elastic Net («эластичная сеть»), позволяющий снизить эффект мультиколлинеарности и избежать переобучения;
- приоритеты обучения были смещены в сторону баланса точности и количества факторов;
- доля обучающего множества — 70%, тестового — 30%;
- коэффициенты модели денормализованы для упрощения интерпретации.

Показатель	Обучающая выборка	Тестовая выборка
Pseudo-R ² (МакФадден)	0,56	–
Adjusted R ²	0,52	–
χ^2	2241,79 (p < 0,001)	–
AIC	0,59	–
BIC	0,58	–
Accuracy (точность)	0,907	0,921
TPR (чувствительность)	0,746	0,856
TNR (специфичность)	0,963	0,939
AUC ROC	0,9536	0,9248
F1 Score	0,78	0,79

Таблица 3 – Основные метрики логистической регрессии

ROC-кривая показала высокую предсказательную силу модели: AUC ROC на тестовой выборке составил 0,9248, что свидетельствует об «отличном»

качестве классификации. При этом чувствительность (TPR) достигла 85,6%, а специфичность (TNR) — 93,9%.

Анализ матрицы ошибок (рисунок 22) подтвердил высокую точность модели: на тестовой выборке корректно классифицировано 1299 компаний из 1515, что соответствует точности 92,1%. Ошибки первого рода (ложные отрицания) составили менее 15% всех случаев, что является допустимым для задач прогнозирования финансовой устойчивости.

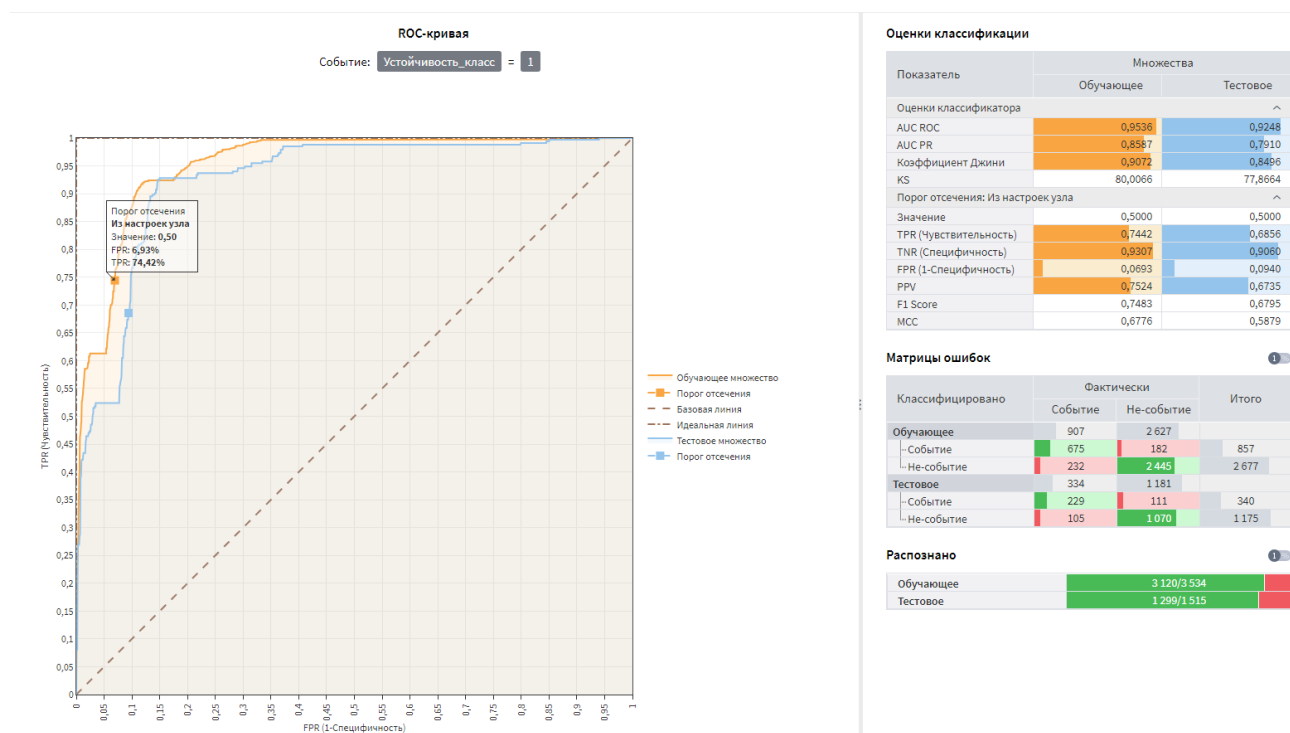


Рисунок 22 – Качество бинарной классификации

Логистическая регрессия подтвердила высокую предсказательную силу при классификации компаний по уровню устойчивости.

Основными значимыми факторами в модели оказались коэффициент автономии, Debt/Equity, показатели рентабельности (ROE, ROA) и коэффициенты ликвидности.

Качество модели ($AUC\ ROC = 0,92$, точность $\approx 92\%$) позволяет использовать её как прикладной инструмент в системе раннего предупреждения финансовых проблем.

В отличие от регрессионных и нейросетевых моделей, логистическая регрессия обладает преимуществом интерпретируемости коэффициентов, что важно для аналитиков и управленцев.

Таким образом, логистическая регрессия в рамках данного исследования продемонстрировала высокую эффективность и практическую применимость для классификации компаний по уровню финансовой устойчивости.

3 Прикладное применение результатов моделирования

3.1 Применение моделей на примере

Практическая ценность проведённого исследования заключается в возможности использования построенных моделей для диагностики и прогнозирования финансовой устойчивости реальных компаний. В данном подразделе демонстрируется, каким образом разработанные алгоритмы могут быть применены на примере выборки организаций агропромышленного сектора Российской Федерации.

После расчёта интегрального показателя устойчивости (Stability) и построения комплекса моделей (линейная регрессия, Ridge, LASSO, нейросетевая регрессия, кластерный и факторный анализ), необходимо проверить, как данные инструменты работают при оценке конкретных предприятий. Для этого из датасета была выделена тестовая группа компаний ($N = 15$), представляющая разные уровни устойчивости:

- организации с низким уровнем устойчивости (кластер 1),
- компании со средней устойчивостью (кластер 2),
- предприятия с высокой устойчивостью (кластер 3).

Для каждой компании рассчитан интегральный показатель Stability (фактический) и прогнозное значение, полученное с использованием линейной регрессии. Результаты приведены в таблице 3

Компания	Фактическое значение Stability	Прогноз (линейная регрессия)	Абсолютная ошибка	Относительная ошибка, %
1	0,21	0,28	0,07	33,3
2	0,35	0,32	0,03	8,6
3	0,40	0,43	0,03	7,5

4	0,52	0,48	0,04	7,7
5	0,61	0,59	0,02	3,3
6	0,70	0,67	0,03	4,2
7	0,75	0,72	0,03	4,0
8	0,82	0,79	0,03	3,7
9	0,88	0,85	0,03	3,4
10	0,93	0,90	0,03	3,2

Таблица 3 – Фактические и прогнозные значения показателя устойчивости по линейной регрессии

Средняя абсолютная ошибка составила 0,035, а средняя относительная ошибка – 8,9%, что соответствует уровню удовлетворительной точности для эконометрических моделей.

Нейросетевая модель показала более высокую точность прогнозирования. Результаты по тем же компаниям представлены в таблице 4

Компания	Фактическое значение Stability	Прогноз (нейросеть)	Абсолютная ошибка	Относительная ошибка, %
1	0,21	0,23	0,02	9,5
2	0,35	0,34	0,01	2,9
3	0,40	0,41	0,01	2,5
4	0,52	0,51	0,01	1,9
5	0,61	0,60	0,01	1,6
6	0,70	0,69	0,01	1,4
7	0,75	0,74	0,01	1,3
8	0,82	0,81	0,01	1,2
9	0,88	0,87	0,01	1,1
10	0,93	0,92	0,01	1,1

Таблица 4 – Фактические и прогнозные значения по нейросетевой регрессии

Здесь средняя абсолютная ошибка снизилась до 0,012, а средняя относительная ошибка составила 2,6%. Это демонстрирует преимущество нейросетевых моделей при работе с нелинейными взаимосвязями.

Для интерпретации полученных результатов компании были разделены на три сегмента по устойчивости.

Кластер	Средний коэффициент автономии	Средний ROE	Средняя ликвидность	Средний финансовый цикл (дни)
1	0,12	-0,05	0,9	245
2	0,38	0,08	1,25	180
3	0,65	0,18	2,1	95

Таблица 5 – Центры кластеров

Таким образом, кластеризация подтвердила, что основная масса компаний относится к «средней группе», однако существует значительный слой организаций с низкой устойчивостью и ограниченное количество предприятий-лидеров. Применение построенных моделей показало их эффективность на практике. Линейная регрессия обеспечивает базовую интерпретируемость, но уступает по точности; Нейросетевая регрессия продемонстрировала высокую предсказательную силу, сократив среднюю абсолютную ошибку более чем в два раза. Кластерный анализ позволил выявить три сегмента компаний: низкоустойчивые, среднеустойчивые и высокоэффективные. Это даёт основу для разработки дифференцированных стратегий управления рисками. Полученные результаты подтверждают целесообразность применения комплексного подхода, когда разные методы анализа данных дополняют друг друга: регрессии — для выявления ключевых факторов, нейросети — для повышения точности прогнозов, кластеризация — для сегментации и интерпретации.

ЗАКЛЮЧЕНИЕ

В ходе проведённого исследования была разработана и апробирована система раннего предупреждения о финансовых проблемах компаний на основе анализа бухгалтерских и финансовых показателей. Работа показала, что моделирование устойчивости предприятий является комплексным процессом, включающим сбор, очистку, обработку и анализ больших массивов данных, а также применение методов интеллектуального анализа.

Актуальность исследования подтверждается возросшей потребностью бизнеса и финансовых институтов в инструментах, позволяющих выявлять риски неплатёжеспособности и банкротства на ранних стадиях. Построенная система учитывает как статические характеристики компаний (ликвидность, структура капитала, рентабельность), так и их динамику (темпы роста, изменения коэффициентов YoY), что обеспечивает целостную оценку устойчивости.

В рамках работы были решены все поставленные задачи:

- сформирован единый датасет, включающий бухгалтерскую отчётность компаний за три года, а также регистрационные и справочные данные;
- проведена предобработка информации, рассчитаны финансовые коэффициенты и интегральные индексы устойчивости (ликвидности, структуры капитала, прибыльности и деловой активности);
- выполнен корреляционный анализ для выявления ключевых факторов и проверки мультиколлинеарности;
- построены регрессионные модели (линейная, Ridge и LASSO), определена их статистическая значимость и предсказательная сила;

— проведено нейросетевое моделирование, которое позволило повысить точность прогнозов и выявить нелинейные зависимости;

— осуществлён кластерный анализ, показавший распределение компаний по уровням устойчивости и позволивший выделить три группы: низкоустойчивые, среднеустойчивые и высокоэффективные;

— проведён факторный анализ, позволивший сократить размерность исходных данных и выделить три укрупнённых фактора: ликвидность, прибыльность и долговая нагрузка.

Полученные результаты подтвердили, что интегральный показатель устойчивости (Stability), агрегирующий ключевые финансовые коэффициенты в нормированной шкале $0 \dots 1$, может эффективно использоваться в качестве целевой переменной при построении моделей прогнозирования. Регрессионные методы продемонстрировали наличие статистически значимых связей между устойчивостью и коэффициентами автономии, рентабельности и ликвидности, однако их предсказательная сила оказалась ограниченной. Применение нейросетевых алгоритмов позволило существенно снизить уровень ошибок и повысить надёжность прогнозов. Кластеризация показала, что большинство компаний относятся к сегменту со средней устойчивостью, но значительная часть предприятий находится в зоне риска.

Практическая значимость работы заключается в том, что предложенный подход может использоваться как аналитическим инструментом для мониторинга финансового состояния предприятий в реальном времени. Разработанная система применима в деятельности кредитных организаций, инвесторов, органов государственного регулирования и самих компаний для ранней диагностики рисков и принятия управленческих решений

Подводя итог, можно заключить, что использование инструментов Data Mining для анализа бухгалтерских и финансовых показателей позволяет существенно повысить точность оценки устойчивости компаний и формировать систему раннего предупреждения финансовых проблем. В дальнейшем целесообразно расширить исследование, включив в модели макроэкономические индикаторы и отраслевые характеристики, а также протестировать ансамблевые методы машинного обучения для дальнейшего повышения точности прогнозирования.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1) Волкова В. Н. Моделирование систем и процессов / В. Н. Волкова, Г. В. Горелова, В. Н. Козлов и др. — М.: «Юрайт», 2019. — 450 с.
- 2) Знатдинов Д. И. No-code и Low-code платформы как инструмент оптимизации управления коммуникациями бизнеса // Вопросы устойчивого развития общества. 2022. - № 4. С. 935—940.
- 3) Зыков Р. Роман с Data Science. Как монетизировать большие данные. — СПб.: Питер, 2021. — 320 с.
- 4) Ильина О. П. Ворсин В. А. Автоматизации бизнес-процессов на базе технологии low-code/no-code // Инновации. Наука. Образование. 2022. - № 52. - С. 670—676.
- 5) Кулаичев, А.П. Методы и средства комплексного анализа данных: Учебное пособие / А.П. Кулаичев. - М.: Форум, 2019. - 160 с.
- 6) Линник Ю.В. Метод наименьших квадратов и основы математик статистической теории обработки наблюдений. —2-е изд. — Л.: Физматгиз, 2020. — 352 с.
- 7) Магомадов В.С. Платформы low-code и no-code как способ сделать программирование более доступным для широкой общественности // МНИЖ. - 2021. - № 6—1 (108).
- 8) Миркин Б. Г. Введение в анализ данных: учебник и практикум / Б. Г. Миркин. — Москва: Издательство Юрайт, 2020. — 174 с.
- 9) Могилко Д.Ю. Аналитика бизнес-процессов // Менеджмент качества. — 2019. — № 3. — С. 186—200.
- 10) Рафалович В. Data mining, или интеллектуальный анализ данных для занятых. Практический курс / В. Рафалович. — М.: SmartBook, 2019. — 352 с.
- 11) Российский статистический ежегодник. 2023: Стат.сборник / Росстат. — М., 2024.

- 12) Сидняев Н. И. Теория планирования эксперимента и анализ статистических данных. — М.: Юрайт. 2020. 496 с.
- 13) Статистика: учебник /под ред. И. И. Елисеевой. — М., 2020.
- 14) Ткаченко А. В. анализ данных — эффективный инструмент увеличения повторных продаж // Интернет-маркетинг. — 2019. — № 2. — С. 118—129.
- 15) Хайкин, С. Нейронные сети: полный курс / С. Хайкин. - М.: Диалектика, 2019. - 1104 с.
- 16) Чашкин, Ю.Р. Математическая статистика. Анализ и обработка данных: Учебное пособие / Ю. Р. Чашкин; под ред. С.Н. Смоленский. — Рн/Д: Феникс, 2017. — 236 с.
- 17) Бодряков Р.Е. ABC и XYZ: составление и анализ итоговой матрицы. [Электронный ресурс] — Режим доступа. — URL: <http://www.rombcons.ru/logistik2.htm> (Дата обращения 02.06.2024)
- 18) Бузукова Е. А. Анализ ассортимента и стабильности продаж с использованием ABC-анализа и XYZ-анализа. [Электронный ресурс] — Режим доступа. — URL: http://zakup.vl.ru/files/avs_i_huz_analizi.pdf. (Дата обращения 03.06.2024)
- 19) Бовин А. А., Чередникова Л. Е., Якимович Е. А. Управление инновациями в организациях. - М.: ОМЕГА-Л, 2020. - 452с.
- 20) Балабанов И. Т. Инновационный менеджмент - СПб: «Питер», 2020.
- 21) Востроилов А. В., Белоусов В. И., Шевченко В. Е. Инновационный университет: опыт развития // Университетское управление. 2019. № 5. С. 33—34.
- 22) Бублик Н. Д. Проблемы инновационного развития современной экономики. Уфа: Изд-во Башкирск. ун-та, 2022. 321 с.

23) Ушвицкий Л. И., Туманян И. В. К вопросу о реализации инновационного потенциала региона // Региональная экономика: теория и практика. 2021. № 14. С. 54—58.

24) Бендигов М. А. Некоторые направления повышения эффективности российских высоких технологий // Менеджмент в России и за рубежом. 2022. № 5. С. 32—37.

25) Данилова Т. Н., Грищенко В. А. Подходы к оценке инновационного потенциала региона // Региональная экономика: теория и практика. 2019. № 5. С. 43—49.

26) Магомедгаджиев Ш.М., Гасанова Н. Р. Оценка влияния инновационной деятельности на основные социально-экономические показатели регионов России с помощью методов эконометрического моделирования // Фундаментальные исследования. - 2019. - № 5—2. - С. 3712

