



Московский государственный университет имени М. В. Ломоносова

Экономический факультет

КАФЕДРА ЭКОНОМИЧЕСКОЙ ИНФОРМАТИКИ

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

**«МОДЕЛИРОВАНИЕ РИСКОВ В БАНКОВСКИХ СИСТЕМАХ С
ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ»**

Работу выполнил

студент группы Э405:

Колесников Дмитрий Михайлович

Научный руководитель:

к.э.н., доцент Скрипкин Кирилл Георгиевич

Москва

2025

Оглавление

ВВЕДЕНИЕ	2
ГЛАВА 1: ТЕОРЕТИЧЕСКИЕ ОСНОВЫ УПРАВЛЕНИЯ БАНКОВСКИМИ РИСКАМИ	5
1.1 Понятие и классификация банковских рисков.....	5
1.2 Традиционные методы моделирования и управления рисками.....	13
1.3 Регуляторные аспекты в управлении банковскими рисками	24
1.4 Роль эконометрики в банковском риск-менеджменте	26
ГЛАВА 2: МАШИННОЕ ОБУЧЕНИЕ И МНОГОАГЕНТНЫЕ МОДЕЛИ В УПРАВЛЕНИИ БАНКОВСКИМИ РИСКАМИ	29
2.1 Обзор методов машинного обучения, применяемых в банковской сфере	29
2.2 Применение методов машинного обучения для моделирования кредитных рисков	38
2.3 Введение в многоагентное моделирование банковских рисков	45
ГЛАВА 3: ЭКОНОМИЧЕСКИЙ ЦИКЛ И КРИЗИСНЫЕ СИТУАЦИИ	51
3.1 Фазы экономического цикла: рост, пик, спад, дно и восстановление	51
3.2 Финансовый кризис 2008 года.....	56
3.3 Последствия для банковского сектора США	58
3.4 Последствия для российского банковского сектора	59
3.5 Выводы после кризиса	61
ГЛАВА 4: ПРОГНОЗИРОВАНИЕ БАНКОВСКОГО ДЕФОЛТА. ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ VS СЛУЧАЙНЫЙ ЛЕС	63
4.1 Описание переменных	63
4.2 Выбор модели и гипотеза	65
4.3 Процесс и результаты моделирования	66
4.4 Оценка моделей и сравнение производительности	73
4.5 Классификация уровней риска	74
ГЛАВА 5: ЭФФЕКТИВНОСТЬ МОДЕЛИРОВАНИЯ НА РАЗНЫХ ЭТАПАХ ЭКОНОМИЧЕСКОГО ЦИКЛА	76
5.1 Фаза роста	76
5.2 Фаза пика.....	79
5.3 Фаза рецессии.....	83
5.4 Фаза дна	86
5.5 Сравнение этапов	90
ЗАКЛЮЧЕНИЕ	94
СПИСОК ЛИТЕРАТУРЫ	96

Введение

Стабильность банковского сектора по-прежнему вызывает серьезную озабоченность в условиях повторяющихся финансовых кризисов и стремительных технологических изменений. Недавние события, в частности банкротства банков в США в марте — мае 2023 года, представляют собой значительный общесистемный банковский стресс со времен Великого финансового кризиса, когда несколько крупных учреждений рухнули или были спасены в течение нескольких дней. Подобные эпизоды подчеркивают сохраняющийся глобальный риск в банковской сфере и высокую вероятность неплатежей по кредитам. В то же время достижения в области анализа данных и искусственного интеллекта меняют подходы к моделированию рисков: регулирующие органы теперь признают, что искусственный интеллект может повысить эффективность принятия кредитных решений при борьбе с мошенничеством и отмыванием денег. В данной работе рассматриваются эти взаимосвязанные тенденции, исследуется, как современные методы машинного обучения (ML) могут улучшить оценку банковских рисков. Исследование продиктовано необходимостью разработки более точных моделей управления рисками для повышения финансовой стабильности. Целью данного исследования является разработка и сравнение количественных моделей риска банковской системы с использованием современных методов управления рисками наряду с традиционными эконометрическими подходами. Основными задачами являются:

1. Разработка эконометрической базовой модели для оценки кредитного риска;
2. Реализация алгоритмов классификации, таких как логистическая регрессия и случайные леса;
3. Проверка и сравнение эффективности модели с реальными банковскими данными;
4. Анализ последствий в рамках нормативно-правовой базы.

Объединив эти элементы, дипломная работа направлена на выявление наиболее эффективных методов прогнозирования дефолта и системного риска в соответствии с требованиями Базеля III и Центрального банка России. Объектом исследования является динамика кредитных рисков банковской системы. Предметом исследования является моделирование этих рисков с использованием передовых количественных методов.

Другими словами, я исследую процессы банковского дефолта (объект) с помощью разных моделей (субъект).

В данной работе используется многосторонний методологический подход:

1. Эконометрическое моделирование - я начал с классических методов, таких как модели логистической регрессии. Эти эконометрические модели оценивают риск дефолта как функцию заемщика и макропеременных. Как отмечается в литературе, логистическая регрессия является стандартом для оценки кредитоспособности.
2. Логистическая регрессия - бинарная логит-модель используется для классификации результатов кредитования (дефолт и не-дефолт). Это простой вероятностный классификатор, который широко используется в исследованиях кредитных рисков.
3. Дерево принятия решений – модель ML, которая разделяет данные по переменным для прогнозирования дефолта. Она интуитивно понятна и может учитывать нелинейность. Множество таких деревьев являются основой для модели случайного леса (RF).

В дипломной работе рассматриваются несколько ключевых вопросов:

1. Могут ли алгоритмы ML (деревья решений, случайные леса) значительно превосходить традиционные логит-модели в прогнозировании дефолтов?
2. Какие переменные наиболее сильно влияют на риск дефолта в рамках моделей?

В эмпирическом анализе используются подробные данные об уровне кредитования и банковских операциях, дополненные экономической статистикой. Макроэкономические переменные (рост ВВП, безработица, инфляция) и рыночные показатели (процентные ставки) взяты из официальных источников (например, Центрального банка России, Росстата. Кроме того, для установки параметров модели и подтверждения коэффициентов достаточности капитала используются нормативные документы и опубликованная финансовая отчетность (банковские отчеты). Все данные обрабатываются и очищаются для обеспечения согласованности (решения таких проблем, как дисбаланс классов) перед оценкой модели.

Эта работа развивает как теорию, так и практику в области моделирования банковских рисков. Теоретически, она расширяет возможности применения ML в реальном банковском контексте. Сочетание эконометрических методов и методов ML обеспечивает новую гибридную основу для анализа рисков. Эмпирически ожидается, что дипломная работа продемонстрирует повышенную точность прогнозирования: в предыдущих исследованиях отмечалось, что кредитные модели на основе ML могут обеспечить повышенную точность прогнозирования, мониторинг рисков в режиме реального времени и персонализированные решения. На практике результаты имеют очевидные последствия для банков и регулирующих органов. Более совершенные классификаторы рисков могут улучшить принятие кредитных решений и распределение капитала, потенциально снижая процент неплатежей и непредвиденных убытков.

Глава 1: Теоретические основы управления банковскими рисками

1.1 Понятие и классификация банковских рисков

Банки подвергаются различным рискам в своей деятельности. Эффективная классификация рисков имеет важное значение для управления рисками и соблюдения нормативных требований. Международные стандарты, такие как стандарты, установленные Базельским комитетом по банковскому надзору (BCBS) при Банке международных расчетов (BIS), сошлись на основном наборе категорий рисков, которыми банки должны управлять и удерживать капитал. Согласно этим глобальным структурам, основные риски в банковской сфере можно классифицировать как кредитный риск, рыночный риск, операционный риск и риск ликвидности. Эта классификация широко принята регулирующими органами (включая Базельские соглашения) и международными институтами, такими как МВФ и БМР, и охватывает основные финансовые и нефинансовые риски, с которыми сталкиваются банки. Примечательно, что многие другие часто упоминаемые типы рисков (такие как риск процентной ставки, валютный риск, риск соответствия и т. д.) на самом деле являются подкомпонентами или конкретными примерами этих четырех широких категорий. Используя эту классификацию рисков, банки обеспечивают комплексное покрытие рисков и соответствие мировым стандартам.

Тип риска	Определение (Авторитетный источник)	Значение и контекст	Измерение и количественная оценка
Кредитный риск	Вероятность того, что заемщик или контрагент не выполнит свои обязательства.	Суть кредитования; главная причина банкротств банков. Встречается в кредитах, облигациях, деривативах и т. д.	Кредитный анализ, рейтинги; Ожидаемые потери = $PD \times LGD \times EAD$. Капитал через активы, взвешенные по риску (Базель).
Рыночный риск	Риск потерь по балансовым/внебалансовым	Возникает в ходе торговой и инвестиционной	Var (Value at Risk), стресс тестирование

Тип риска	Определение (Авторитетный источник)	Значение и контекст	Измерение и количественная оценка
	ым позициям из-за движения рыночных цен.	деятельности (процентные ставки, валюта, акции, товары)	
Операционный риск	Риск потерь из-за неадекватных или неудачных процессов, людей, систем или внешних событий. (Включает юридический риск; исключает стратегический/репутационный)	Мошенничество, технологические сбои, неправомерные действия, кибератаки). Затрагивают все виды коммерческой деятельности.	Внутренний контроль, аудит, KRI, анализ сценариев. Капитал по стандартизованным формулам (например, 15% от валового дохода по базовому индикаторному подходу).
Риск ликвидности	Риск того, что банк не сможет выполнить обязательства в срок, не понеся убытков. Включает: риск ликвидности финансирования (неспособность удовлетворить потребности в денежных потоках) и риск ликвидности рынка (неспособность продать активы по справедливой цене).	Критически важно из-за трансформации сроков погашения банков. Проявляется в банковских набегах или заморозках рынка. Краткосрочные (денежный поток) и долгосрочные (финансирование) аспекты.	Метрики покрытия ликвидности (LCR, NSFR), стресс-тесты. $LCR = HQLA / 30\text{-дневный чистый отток} \geq 100\%$ (требование Базеля III).

Таблица 1. Основные категории банковских рисков, их определения, важность и измерение. «PD» = Вероятность дефолта; «LGD» = Убыток при дефолте; «EAD» = Подверженность риску при дефолте; «HQLA» = Высококачественные ликвидные активы; «LCR» = Коэффициент покрытия ликвидности.

Кредитный риск измеряет потенциальные потери в случае невыполнения заемщиками или контрагентами своих обязательств. Основопологающим показателем является ожидаемый убыток (EL), рассчитываемый как произведение вероятности дефолта (PD), убытка в случае дефолта (LGD) и риска в случае дефолта (EAD). В виде формулы:

$$EL = PD \cdot LGD \cdot EAD$$

который представляет собой средний убыток, ожидаемый банком от риска. Например, если кредит имеет 2% PD, 40% LGD и риск в размере 1 миллиона долларов, ожидаемый убыток равен \$8,000 ($0.02 \times 0.40 \times \$1,000,000$). Банки также должны учитывать непредвиденные убытки (UL) – потенциальные убытки, превышающие ожидаемый уровень, которые обычно учитываются в требованиях к капиталу. В соответствии с базельским подходом, основанным на внутренних рейтингах (IRB), регулятивный капитал должен покрывать UL с высокой степенью достоверности (99,9%).

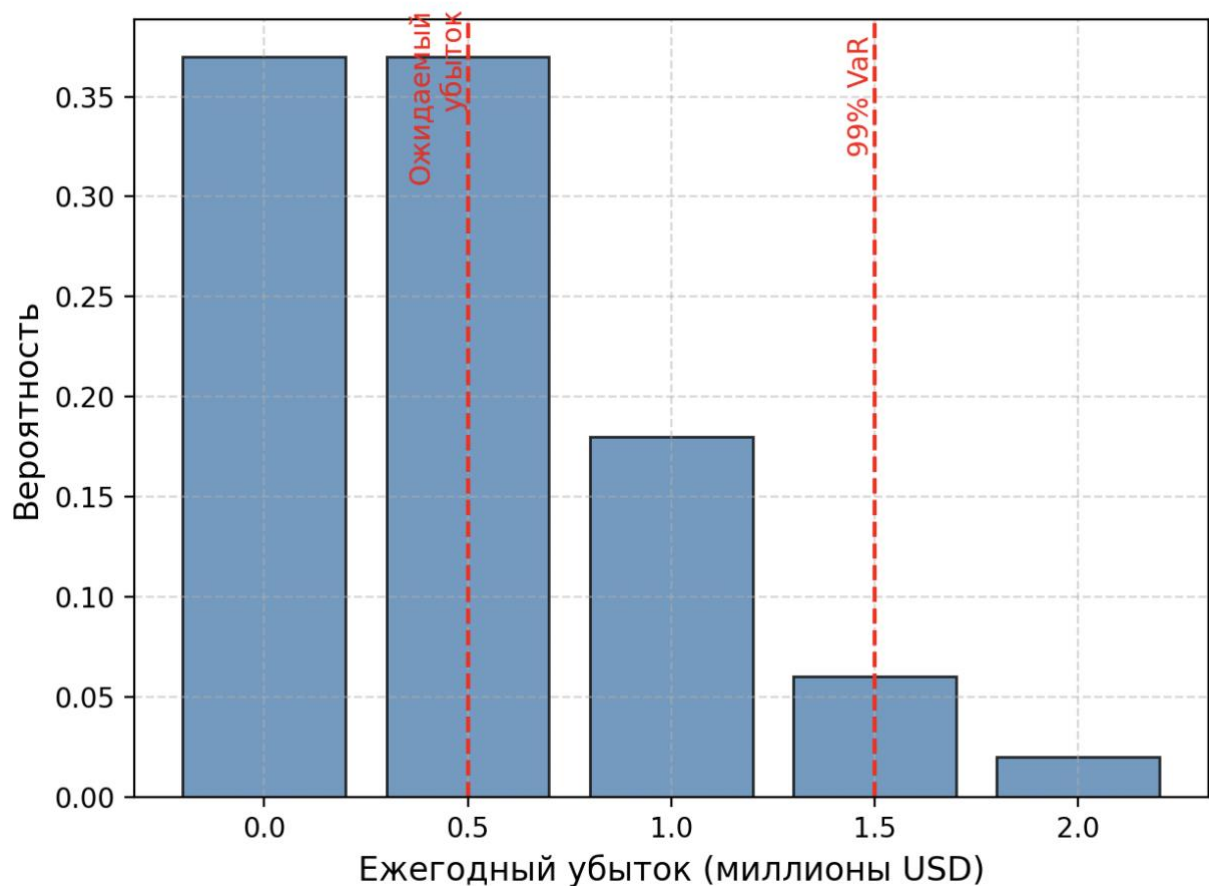


Рисунок 1. Моделируемое распределение годовых кредитных убытков для портфеля (100 кредитов с 1% PD и 40% LGD каждый). Дискретная масса вероятности (вероятность того, что дискретная случайная величина примет фиксированное значение) показывает наибольшую вероятность при нулевом или одном дефолте. Ожидаемый убыток (EL) отмечен на уровне 0,4 миллиона долларов США, а 99% VaR (99-й процентиль убытка) - на уровне ~1,6 миллиона долларов США, что покрывает непредвиденные убытки. Полосы справа от VaR (очень низкая вероятность) представляют собой хвостовые события за пределами 99-процентного уровня. Речь идет о традиционных представлениях об управлении рисками. У нас есть скрытая неопределенность (вроде той, что была в 2007), которую не измеряет VaR.

На рисунке 1 видно, что распределение кредитных убытков сильно асимметрично, с высокой вероятностью нулевых или небольших убытков и длинным правым хвостом для нескольких дефолтов. Ожидаемый убыток (красная пунктирная линия на уровне 0,4) равен $PD \times LGD \times EAD$ (для одного среднего дефолта). Однако существует небольшая вероятность нескольких одновременных дефолтов - неожиданный убыток. 99%-ная стоимость риска (VaR) на уровне 1,6 (четыре дефолта) указывает на 1%-ную вероятность

того, что потери превысят этот уровень. Банки хранят капитал на случай таких экстремальных, но вероятных кредитных потерь. Формула IRB, приведенная выше, дает аналогичный квантиль 99,9% (не показан на графике), используемый для создания резервов под капитал, гарантируя, что банк сможет поглотить очень редкие кластеры дефолтов. На практике регулирующие органы, такие как Базельский комитет, требуют сравнивать общий ожидаемый убыток по всем рискам с резервами и требовать капитал только для непредвиденной части, что соответствует идее, что обычные убытки покрываются ценообразованием/резервами, и только хвостовой риск требует капитала. Рыночный риск — это риск потерь, связанных с изменением рыночных цен (процентных ставок, обменных курсов, цен на акции и т. д.). Ключевой количественной метрикой является *Value at Risk (VaR)*, которая при заданном уровне доверия (например, 99 %) представляет собой порог потерь, который будет превышен только с такой малой вероятностью. Другими словами, 99% *VaR* — это размер убытков, который будет превышен только в 1% случаев. Например, если предполагается, что ежедневные торговые P/L (прибыли и убытки) распределены нормально со стандартным отклонением σ , 99% - ый однодневный *VaR* составляет примерно $2,33\sigma$ (поскольку 2,33 — это 99-й процентиль стандартного нормального распределения). Банки часто масштабируют *VaR* для различных горизонтов - в Базельских правилах 1996 года использовался 10-дневный горизонт для *VaR* торговой книги, равный $VaR_{10\text{ day}} = \sqrt{10} \cdot VaR_{1\text{ day}}$.

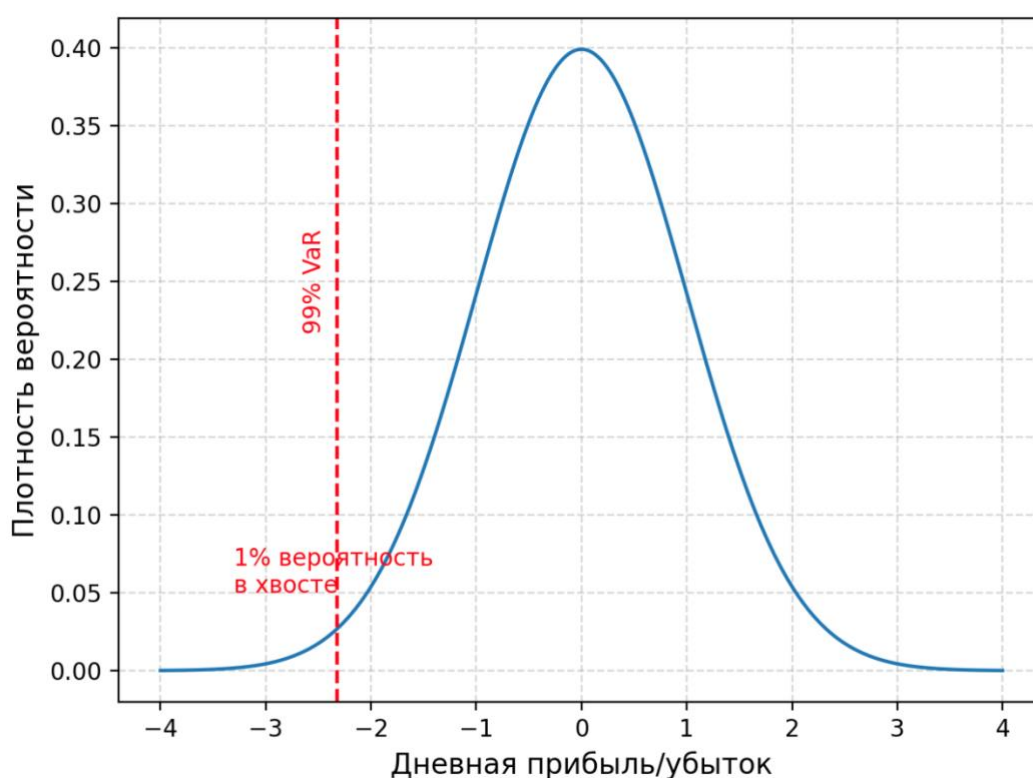


Рисунок 2. Иллюстрация Value-at-Risk (VaR) на примере ежедневного распределения P/L портфеля (предполагается, что оно нормальное).

Кривая показывает плотность вероятности ежедневных прибылей/убытков. 99%-ный VaR обозначен красной пунктирной линией в левом хвосте, так что за пределами этого уровня потерь площадь под кривой составляет 1%. Заштрихованная красная область представляет собой хвост наихудших потерь в 1%.

Операционный риск — это риск потерь в результате сбоев в работе процессов, людей, систем или внешних событий, включая мошенничество, судебные убытки, сбои в работе систем и т. д. В отличие от кредитного и рыночного риска, потери от операционного риска носят нерегулярный характер и часто имеют большой хвост. Банки используют различные подходы к количественной оценке операционного риска. Согласно Базелю II (Basic Indicator Approach, BIA), капитал представляет собой фиксированный процент (α) от валового дохода. Например, для BIA Базель установил $\alpha = 15\%$.

$$KBAI = 0.15 \cdot (\text{Average Annual Gross Income})$$

Эта формула предполагает, что больший доход предполагает большую подверженность операционному риску. Например, банк со средним валовым доходом в 1 миллиард долларов будет иметь капитал под операционный риск в размере 150 миллионов долларов в соответствии с BIA. Более сложные методы были разрешены в рамках усовершенствованного подхода к оценке (Advanced Measurement Approach, AMA) (в настоящее время замененного стандартизированным подходом Базеля III). В соответствии с AMA банки должны были моделировать полное распределение операционных убытков (сочетая частоту и тяжесть событий) и устанавливать капитал для покрытия непредвиденных убытков с вероятностью 99,9% в течение одного года. То есть банки должны были иметь капитал, который превышал бы годовые убытки по риску только в 0,1% случаев.

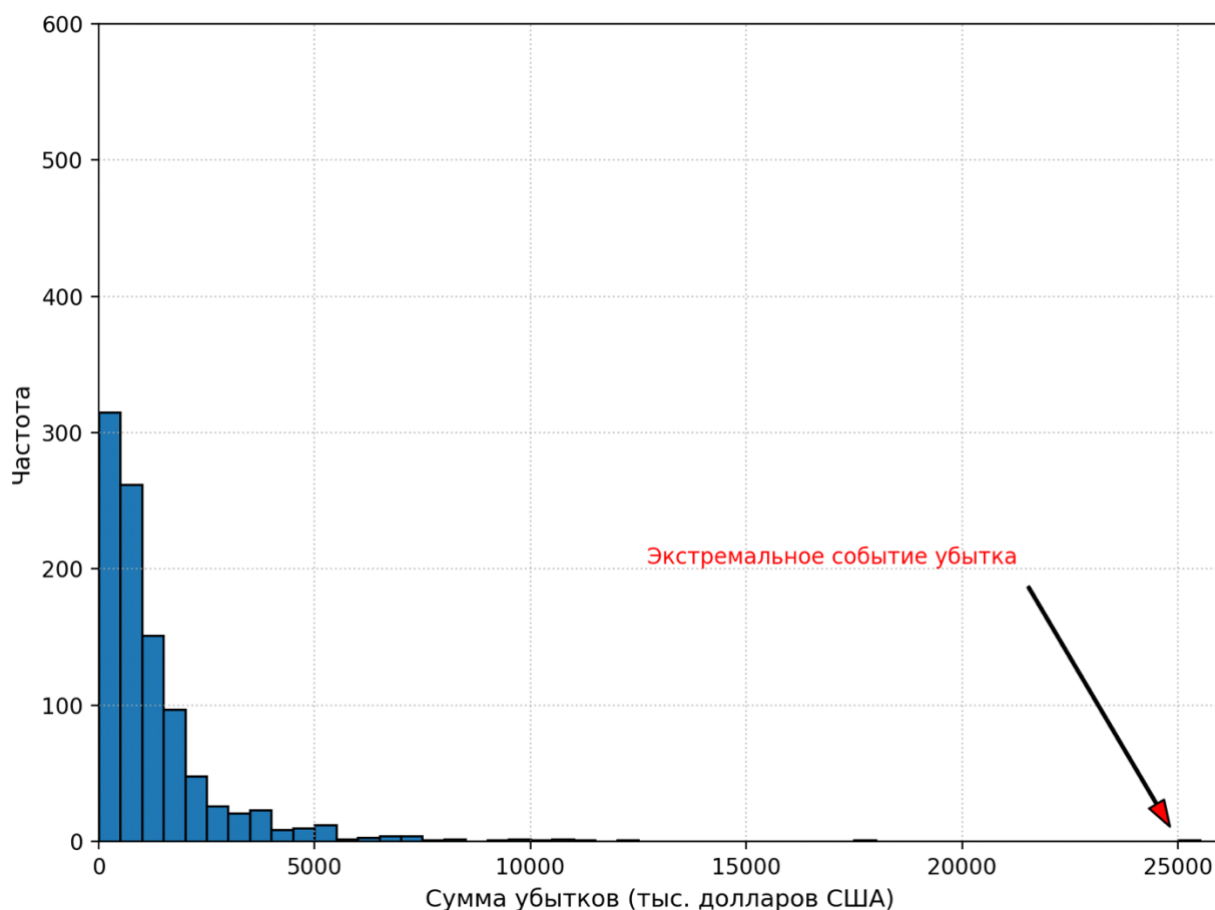


Рисунок 3. Характер операционного риска с большим хвостом.

Подавляющее большинство случаев потерь являются незначительными (например, мелкие мошенничества, ошибки обработки), о чем свидетельствуют высокие столбики слева (менее \$1 000 тыс.). Однако у распределения есть длинный правый хвост: несколько экстремальных событий достигают десятков миллионов (показано одно смоделированное событие стоимостью около 25 миллионов долларов). Эти редкие потери доминируют в квантиле 99,9%. Фактически, максимальный убыток в выборке не соответствует обычной шкале, что подчеркивает, почему традиционные методы VaR не работают - убытки от операционного риска с большим хвостом не поддаются традиционным статистическим выводам, когда все точки данных используются для оценки максимального убытка. Для моделирования таких хвостов банки используют такие методы, как теория экстремальных значений. Современный стандартизированный подход к измерению (SMA) сочетает в себе бизнес-индикатор (показатель риска, основанный на финансовой отчетности) и внутренний опыт банка в отношении убытков для определения капитала. Тем не менее, основной принцип остается неизменным: регуляторы устанавливают для капитала под операционный риск вероятность сбоя в 0,1% в год, что отражает стандарт убытков «тысячелетней давности». В более простом случае правило валового дохода в 15% (BIA) и

его преемники гарантируют, что даже банки, не использующие передовые модели, будут иметь базовый буфер капитала для покрытия операционных убытков.

Риск ликвидности — это риск того, что банк не сможет удовлетворить свои потребности в денежных средствах или залоговом обеспечении в краткосрочной перспективе. Два ключевых регулятивных показателя, введенных в Базель III, — это коэффициент покрытия ликвидности (LCR) и коэффициент стабильного финансирования (NSFR). LCR определяется как:

$$LCR = \frac{\text{Запас высококачественных ликвидных активов}}{\text{Общий чистый отток денежных средств за 30 дней}} \geq 100\%$$

Данная формула гарантирует, что банк располагает достаточным количеством высококачественных ликвидных активов (HQLA) (денежные средства, государственные облигации и т. д.) для покрытия 30-дневного чистого оттока денежных средств в условиях сильного стресса. «Чистый отток рассчитывается путем применения стандартизированных ставок стока к обязательствам и внебалансовым обязательствам за вычетом ожидаемого притока в течение 30 дней. Например, если прогнозируемый 30-дневный отток средств банка составляет \$50 млрд, а приток - \$45 млрд (чистый отток - \$5 млрд), он должен хранить не менее \$5 млрд в HQLA, чтобы $LCR = 100\%$. NSFR, напротив, направлен на обеспечение более долгосрочной ликвидности, требуя наличия стабильного профиля финансирования сроком не менее одного года:

$$NSFR = \frac{\text{Доступное стабильное финансирование (ASF)}}{\text{Требуемое стабильное финансирование (RSF)}} \geq 100\%$$

$NSFR = \text{Доступное стабильное финансирование (ASF)} / \text{Требуемое стабильное финансирование (RSF)} \geq 100\%$

Показатель ASF взвешивается по долгосрочным обязательствам (например, депозитам клиентов, долгосрочному долгу), а показатель RSF взвешивается по активам / балансовым рискам (более высокий вес для менее ликвидных, долгосрочных активов). Формула NSFR гарантирует, что неликвидные активы банка в основном финансируются из стабильных источников, а не из волатильных краткосрочных средств.

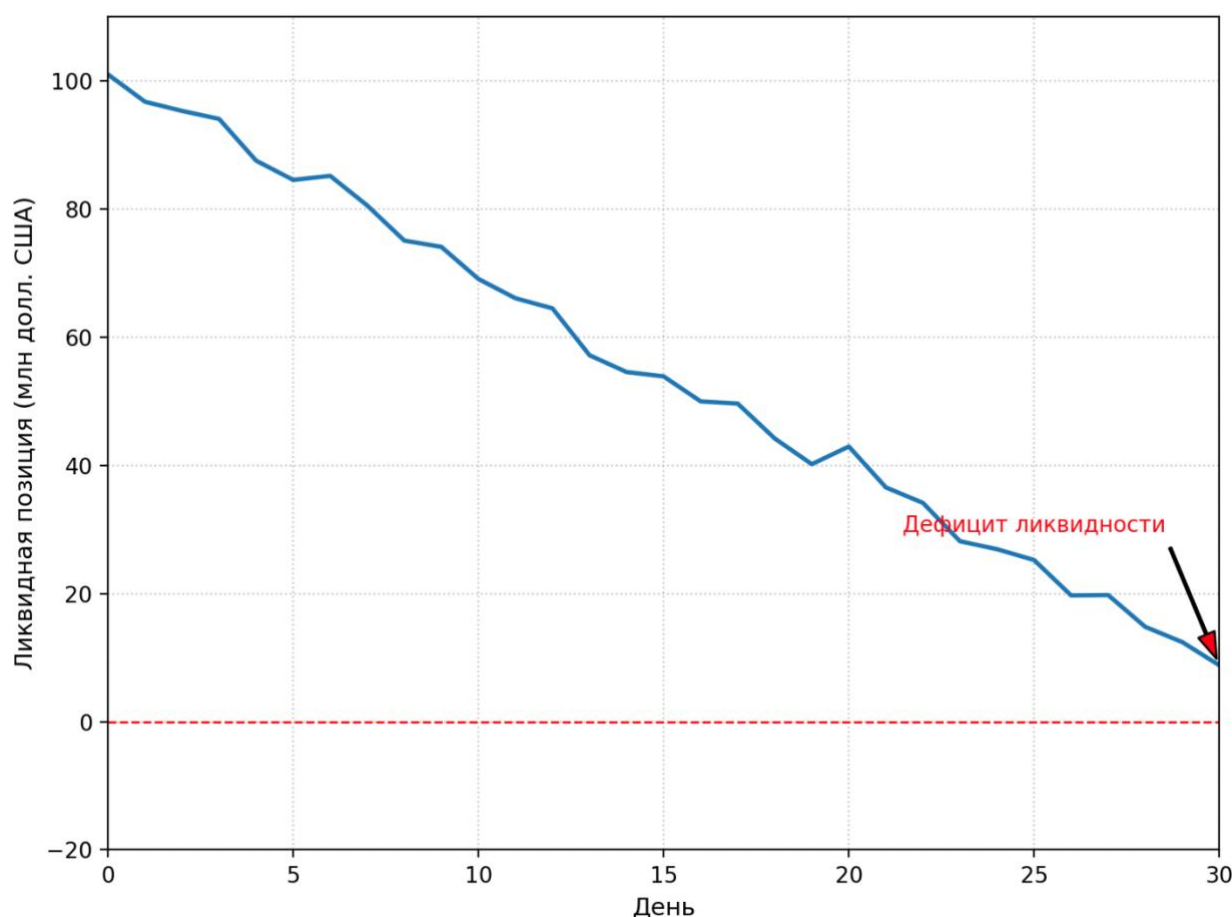


Рисунок 4. Пример разрыва ликвидности за 30 дней.

Синяя линия показывает прогнозируемую позицию банка по ликвидности (остаток денежных средств или HQLA) на каждый день, начиная со 100 (млн долларов США). Красная пунктирная линия — это нулевой уровень. В этом стрессовом сценарии ежедневные оттоки превышают притоки, что приводит к постоянному снижению ликвидности. К 30-му дню ликвидность банка исчерпывается (опускается чуть ниже нуля). Стрелкой показан дефицит ликвидности на 30-й день. Такой профиль используется для оценки 30-дневного горизонта выживания для соответствия LCR.

1.2 Традиционные методы моделирования и управления рисками

Методы управления банковскими рисками — это совокупность способов и инструментов, которые помогают минимизировать финансовые потери. Прежде всего, для эффективной организации управления банковскими рисками необходимо их моделирование. Так, например, модели банковского риска позволяют анализировать данные о кредитной истории, доходах и прочих факторах для прогнозирования событий

дефолта, потерь при дефолте, кредитной конверсии и т. д. Также для моделирования кредитного риска используется имитационное моделирование, которое опирается на статистические распределения параметров заёмщиков в прошлом и учитывает влияние на них различных социально-экономических факторов, используемых в моделировании и управлении банковскими рисками: логистическую регрессию, линейную регрессию, кластеризацию и имитационное моделирование. Каждый из этих методов предлагает свой подход к пониманию и снижению финансовых рисков - от прогнозирования дефолтов по кредитам до сегментирования портфелей и имитации убытков.

Среди традиционных методов моделирования и управления банковскими рисками можно выделить:

1. логистическая регрессия;
2. линейная регрессия;
3. кластеризация;
4. имитационное моделирование.

Логистическая регрессия — это статистический метод, используемый для решения задач бинарной классификации, моделирующий вероятность исхода, который может принимать одно из двух значений (например, дефолт или не дефолт). В отличие от линейной регрессии, которая прогнозирует непрерывный результат, логистическая регрессия прогнозирует логарифмическую вероятность бинарного события с помощью логистической (сигмоидальной) функции. Логистическая модель предполагает, что связь между предикторами и вероятностью события, например дефолт по кредиту задается сигмоидальной функцией:

$$P(Y = 1 | X) = \frac{1}{(1 + \exp [-(\beta_0 + \beta_1 \cdot X_1 + \dots + \beta_p X_p)])}$$

Где $\beta_0 + \beta_1 \cdot X_1 + \dots + \beta_p X_p$ - параметры (коэффициенты) модели, подлежащие оценке. Эти параметры обычно оцениваются методом максимального правдоподобия, максимизируя вероятность наблюдения заданных данных.

Логистическая регрессия уже давно является одним из наиболее широко используемых статистических методов в банковском деле и финансах, который ценится за свою простоту, интерпретируемость (коэффициенты можно интерпретировать как влияние

фактора риска на логарифмический коэффициент дефолта) и хорошо понятные статистические свойства. В банковской сфере логистическая регрессия наиболее известна в моделировании кредитных рисков, особенно для кредитного скоринга и оценки вероятности дефолта (PD). Практически все крупные банки используют модели кредитного скоринга на основе логистической регрессии для принятия решений о выдаче кредитов и установления кредитных лимитов. Например, банк может использовать такие характеристики заемщика, как доход, отношение долга к доходу и кредитная история, в качестве входных переменных в логистической модели, которая выводит вероятность того, что заемщик не выплатит кредит. Поскольку результатом является вероятность (от 0 до 1), логистическая регрессия естественным образом подходит для прогнозирования дефолта и является краеугольным камнем параметров риска Базельского регулятора (PD, потери при дефолте и т. д.). Фактически, логистическая регрессия стала стандартным методом де-факто для кредитного скоринга и считается моделью «белого ящика», поскольку ее решения могут быть объяснены путем изучения весов входных факторов. Помимо кредитного риска, логистическая регрессия также используется для выявления мошенничества и операционного риска (для классификации событий как высокорисковых или нет), и даже в маркетинге (для прогнозирования оттока клиентов или использования продукта) - в любом сценарии, где исход является бинарным.

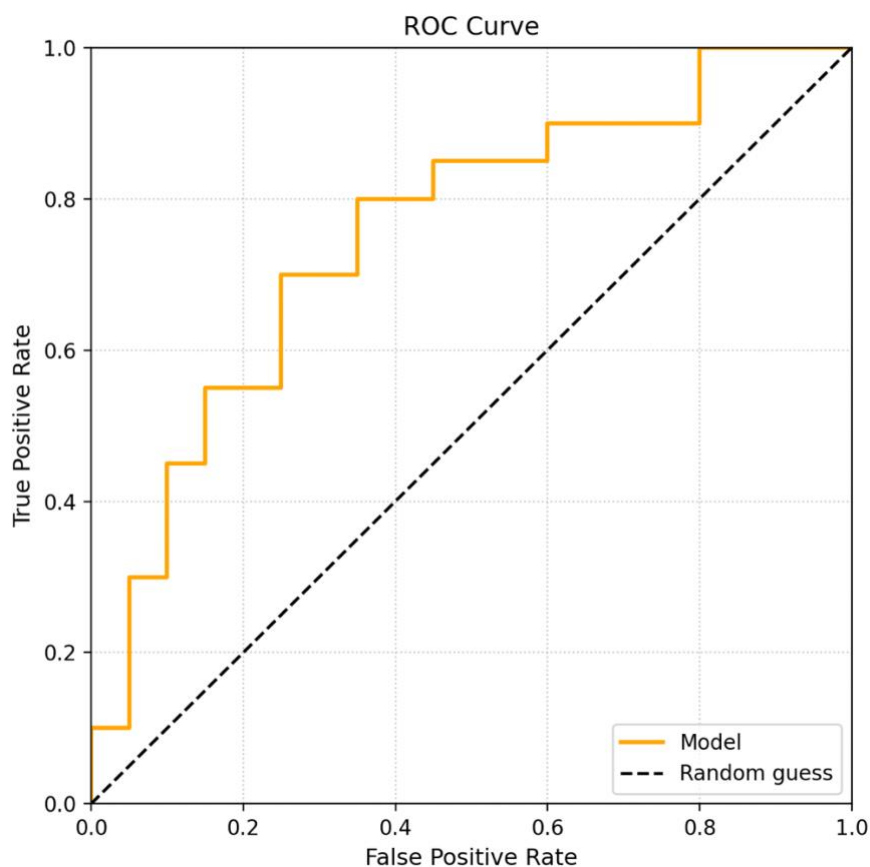


Рисунок 5. ROC – кривая.

Эффективность модели в логистической регрессии обычно оценивается с помощью таких показателей, как кривая операционной характеристики получателя (ROC) и площадь под кривой (AUC). ROC-кривая представляет собой график зависимости частоты истинных положительных результатов (чувствительности) от частоты ложных положительных результатов (1 специфичности) при изменении порога дискриминации.

ROC-кривая на рисунке иллюстрирует, насколько хорошо модель отличает дефолты от не дефолтов при всех порогах классификации - чем ближе оранжевая кривая к левому верхнему углу (высокий процент истинных положительных результатов, низкий процент ложных положительных результатов), тем лучше модель. Диагональная линия представляет собой случайное угадывание. На практике высокий показатель AUC (близкий к 1) указывает на сильную модель. Логистические модели в банковском деле обычно стремятся к высокому значению AUC, чтобы регулирующие органы и риск-менеджеры могли быть уверены в том, что модель упорядочивает риски (например, выявляет наиболее рискованных заемщиков). Выбрав соответствующую границу прогнозируемой вероятности (например, обозначив любой кредит с $PD > 0,5$ как «высокорисковый»), банки могут управлять выдачей кредитов таким образом, чтобы

отклонять большинство плохих кредитов, сводя к минимуму отклонение хороших кредитов. Таким образом, логистическая регрессия представляет собой мощный и в то же время интерпретируемый инструмент для принятия решений о бинарных рисках, который является основой для современного управления кредитными рисками.

Линейная регрессия — это фундаментальный метод моделирования взаимосвязи между непрерывной зависимой переменной и одной или несколькими независимыми переменными (признаками). В своей простейшей форме (простая линейная регрессия с одним предиктором) она проводит прямую линию через набор точек данных, чтобы объяснить, как изменения в предикторе соответствуют изменениям в результате. Общая форма линейной регрессионной модели с предикторами такова:

$$y_i = \beta_0 + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + \dots + \beta_p \cdot x_{ip} + \varepsilon_i, \quad i = 1, \dots, n,$$

где y_i — это ответ (например, метрика риска или финансовый результат для наблюдения), x_{ij} — это значения предикторов, β_0 — это отсекаемый отрезок, β_j — это коэффициенты, а ε_i — это ошибка для наблюдения. Коэффициенты обычно оцениваются с помощью метода наименьших квадратов (МНК), который находит значения β_j , минимизирующие сумму квадратов остатков $\sum_i (y_i - \hat{y}_i)$. Ключевые предположения линейной регрессии включают линейность, независимость ошибок, гомоскедастичность (постоянная дисперсия ошибок) и (часто) нормальность ошибок; когда они выполняются, оценка МНК имеет желаемые свойства, такие как несмещенность и максимальная точность (теорема Гаусса-Маркова).

В банковском деле и финансах линейная регрессия широко используется для моделирования и прогнозирования факторов риска. Она обеспечивает простой способ количественной оценки взаимосвязей между переменными, что делает ее полезной для стресс-тестирования и анализа «что, если». Например, банк может использовать множественную линейную регрессию для связи макроэкономических переменных с убытками по кредитам портфеля, тем самым предсказывая, как изменяются убытки, если растет безработица или падает ВВП. Рассмотрим упрощенный пример оценки кредитного риска, смоделированной как линейная комбинация характеристик заемщика:

$$RiskScore = \beta_0 + \beta_1 \cdot Income + \beta_2 \cdot CreditHistory + \beta_3 \cdot EmploymentDuration + \varepsilon,$$

где $\beta_1, \beta_2, \beta_3$ представляют собой веса дохода, кредитной истории и стажа работы заявителя соответственно. На практике, в то время как логистическая регрессия предпочтительна для бинарных результатов, таких как дефолт (чтобы сохранить прогноз в

[0,1]), линейная регрессия может использоваться для моделирования непрерывных результатов риска, таких как убыток при дефолте (LGD) или подверженность дефолту (EAD), которые являются пропорциями или суммами. Более того, модели линейной регрессии часто лежат в основе прогнозов экономического капитала и финансовых показателей: например, моделирование прибыли банка как линейной функции процентных ставок или прогнозирование резервов на возможные потери по ссудам на основе уровня безработицы. Эти модели ценятся за их интерпретируемость — коэффициент β_j напрямую указывает, как изменение фактора на одну единицу линейно влияет на результат, сохраняя другие факторы постоянными.

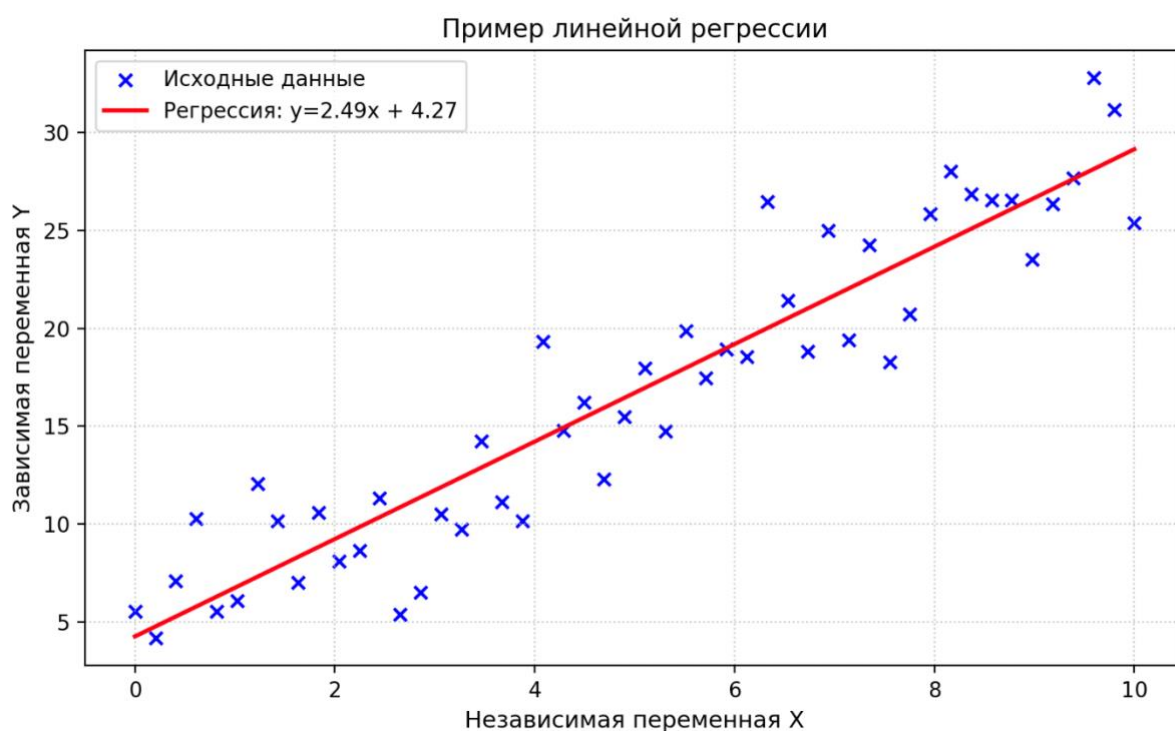


Рисунок 6. Линейная регрессия.

Пример линейной регрессии (красная линия) через точки данных (синие), представляющие связь между независимой переменной X и зависимым результатом Y. На рисунке красная линия — это наилучшая подгонка по методу наименьших квадратов, а синие крестики — это наблюдаемые точки данных. Различия между точками и линией (вертикальные зеленые сегменты на идеальной иллюстрации) — это остаточные ошибки. Меньший разброс точек вокруг линии указывает на более сильную линейную связь (более высокий R^2). В управлении рисками такой график может представлять, например, связь между кредитным плечом заемщика и серьезностью убытков по кредиту или между изменениями процентной ставки и изменениями стоимости портфеля облигаций.

Финансовые аналитики изучают статистику соответствия (например, коэффициент детерминации R^2), чтобы оценить, какая часть изменчивости в результате риска объясняется моделью. Хотя линейная регрессия относительно проста, она остается краеугольным камнем для принятия решений на основе данных в банковской сфере, помогая в оценке кредитного риска, ценообразовании и стресс-тестировании. Правильное применение линейных моделей (с диагностикой предположений) позволяет риск-менеджерам количественно оценивать чувствительность и делать обоснованные прогнозы относительно будущих уровней риска при различных сценариях.

Кластеризация — это метод неконтролируемого обучения, который включает в себя разбиение набора данных (например, клиентов, транзакций, активов) на группы (кластеры) таким образом, что точки данных в одном кластере более похожи друг на друга, чем на точки в других. В отличие от регрессии, кластеризация не имеет предопределенной выходной переменной; вместо этого она стремится обнаружить внутреннюю структуру в данных. Одним из наиболее распространенных алгоритмов кластеризации является кластеризация k -средних, которая делит набор данных на k кластеров путем минимизации внутрикластерной дисперсии. Формально, k -средние находят кластерные назначения

$$C_1, C_2, \dots, C_k$$

и центроиды

$$\mu_1, \mu_2, \dots, \mu_k$$

которые решают задачу оптимизации:

$$\min_{C_1, \dots, C_k, \mu_1, \dots, \mu_k} \sum_{j=1}^k \sum_{x \in C_j} \|x - \mu_j\|^2$$

где $\|x - \mu_j\|^2$ — это евклидово расстояние между точкой x и центроидом μ_j кластера j .

В банковском деле методы кластеризации широко используются для сегментации клиентов, сегментации рисков и обнаружения аномалий. Группируя клиентов в кластеры со схожими характеристиками, банки могут адаптировать стратегии управления рисками и предложения продуктов для каждого сегмента. Например, кластеризация k -средних может сегментировать клиентов кредитных карт на такие кластеры, как «клиенты с

низким уровнем риска и высокой стоимостью» и «клиенты с высоким уровнем риска и нерегулярными расходами», что позволяет дифференцировать кредитные лимиты и мониторинг. Кластеризация может также применяться к кредитным портфелям: кредиты могут быть сгруппированы по атрибутам (отрасль, география, качество кредита) для выявления концентрации риска. При моделировании кредитного риска подход заключается в том, чтобы сначала выполнить кластеризацию, а затем построить отдельные модели *PD* или потерь для каждого однородного кластера заемщиков, что может улучшить предсказательную силу, фиксируя поведение, специфичное для сегмента. Другое важное применение — обнаружение мошенничества и борьба с отмыванием денег (*AML*): кластеризация данных о транзакциях может помочь выявить необычные кластеры транзакций, которые отклоняются от обычных шаблонов, отмечая потенциальную мошенническую деятельность или подозрительное поведение. Поскольку кластеризация не контролируется, ее часто используют в качестве исследовательского инструмента для обнаружения скрытых структур, например, для обнаружения того, что определенные комбинации транзакционных характеристик образуют отдельную группу, которая коррелирует с более высоким операционным риском.

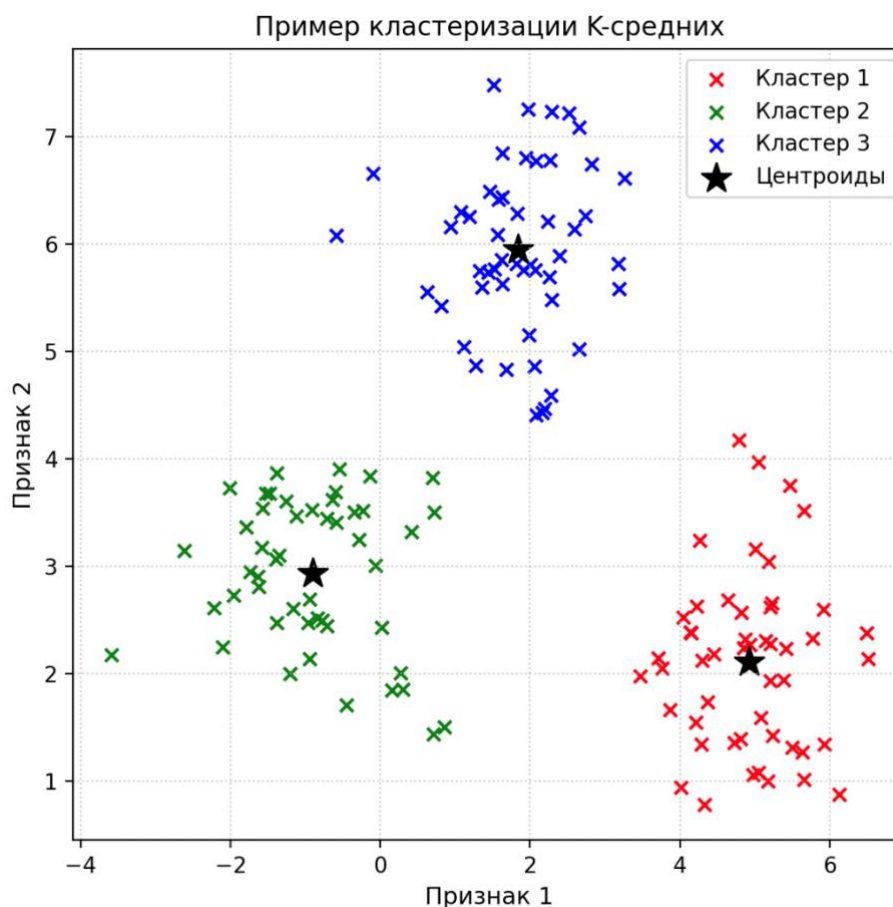


Рисунок 7. Кластеризация K – средних.

Пример кластеризации k-средних на наборе точек данных в двух измерениях. На этом рисунке точки данных были разделены на три кластера (окрашены в красный, зеленый и синий цвета), а центроиды кластеров отмечены черными символами «X». Мы видим, что точки в каждой цветной группе лежат близко к своему центроиду, иллюстрируя идею внутрикластерного сходства. В банковском контексте каждый кластер может представлять отдельный сегмент: например, красные точки могут быть потребительскими кредитами со средним размером и средним риском, зеленые точки могут быть кредитами малому бизнесу с меньшими суммами, но более высокими ставками дефолта, а синие точки могут быть крупными корпоративными кредитами с большими суммами, но более низким относительным риском. Анализируя такие кластеры, риск-менеджеры получают представление о характеристиках каждой группы (например, среднее отношение кредита к стоимости, общие ковариаты) и могут управлять каждым кластером соответствующим образом — возможно, распределяя капитал или устанавливая лимиты риска для каждого кластера. Таким образом, кластеризация способствует диверсификации портфеля (гарантируя, что все риски не попадут в один рискованный кластер) и помогает адаптировать меры по снижению риска (например, улучшенный мониторинг кластера, определенного как наиболее рискованный). В целом, кластеризация обеспечивает способ управления сложностью путем разбиения большого неоднородного портфеля на управляемые подгруппы с общими чертами, что бесценно для управления кредитным портфелем и профилирования риска клиентов.

Имитационное моделирование, в частности моделирование Монте-Карло, является мощным методом оценки банковских рисков путем моделирования неопределенности результатов. Моделирование Монте-Карло использует случайную выборку для аппроксимации вероятностных распределений потенциальных результатов в финансовой модели. Вместо того, чтобы полагаться на одну «лучшую догадку» или несколько детерминированных сценариев, методы Монте-Карло генерируют большое количество моделируемых сценариев для факторов риска (например, процентные ставки, события дефолта, рыночная доходность), а затем вычисляют результирующий убыток портфеля или другую метрику риска для каждого сценария. Объединяя эти результаты, можно оценить полное распределение убытков и вывести меры риска, такие как стоимость под риском (VaR) или ожидаемый дефицит с заданным уровнем достоверности.

Технически, моделирование Монте-Карло включает в себя выборку N независимых выборок случайных входов и вычисление результата для каждой из них. Например, если L — это случайный убыток портфеля, мы генерируем $L^{(1)}, L^{(2)}, \dots, L^{(N)}$ путем случайной выборки дефолтов, движений рынка и т. д. Затем ожидаемый убыток можно оценить с помощью выборочного среднего:

$$\hat{E}[L] = \frac{1}{N} \cdot \sum_{i=1}^N L^{(i)}$$

и высокопроцентный убыток (VaR) можно оценить, взяв $(1 - \alpha)$ квантиль отсортированных убытков. По закону больших чисел эти оценки сходятся к истинным метрикам риска, когда N становится большим. Моделирование Монте-Карло помогает включить риск и неопределенность таким образом, как статические модели не могут. Они требуют предположений о стохастических процессах, определяющих факторы риска (например, вероятности дефолта, волатильность процентных ставок, корреляции между активами), и качество моделирования зависит от обоснованности этих предположений.

На практике моделирование Монте-Карло повсеместно применяется в управлении рыночным риском и риском кредитного портфеля. Известным применением является вычисление стоимости под риском для торговых портфелей: модель Монте-Карло VaR переоценивает портфель в соответствии с тысячами случайных рыночных сценариев (шоки цен, ставок и т. д.), а затем определяет порог убытков, который превышает в 5% сценариев (это 95% VaR). Фактически, Монте-Карло VaR часто рассматривается как золотой стандарт для измерения риска из-за его гибкости в работе со сложными портфелями и нелинейными инструментами. Для кредитного риска банки используют моделирование для моделирования распределения убытков по кредитам в кредитном портфеле путем моделирования событий дефолта и возмещений для каждого заемщика в ходе множества испытаний. Этот подход охватывает эффекты портфеля, такие как корреляция дефолта и риск концентрации, которые трудно оценить аналитически. Регулирование операционного риска (Базель II/III) также использует метод Монте-Карло в подходе распределения убытков, где банки моделируют годовые общие убытки на основе распределения частоты и серьезности событий операционных убытков, а затем вычисляют 99,9% процентиля убытков в качестве требования к капиталу. Помимо количественной оценки риска, моделирование используется в стресс-тестировании и анализе сценариев: например, для оценки вероятности того, что учреждение истощит свой капитал в рамках

ряда моделируемых сценариев рецессии, или для прогнозирования потребностей в ликвидности в рамках сценариев случайного оттока депозитов.



Рисунок 8. Распределение потерь.

Пример смоделированного распределения убытков для гипотетического портфеля, полученного с помощью моделирования Монте-Карло. Гистограмма выше показывает частоту результатов убытков портфеля (по оси x) из многих испытаний моделирования. Мы видим, что большинство смоделированных результатов сгруппированы в сторону более низких убытков (пик слева), в то время как длинный правый хвост указывает на небольшую вероятность очень больших убытков. Это скошенное распределение убытков характерно для кредитного и операционного риска, где убытки случаются нечасто, но могут быть серьезными. Из такого распределения менеджеры по рискам могут извлечь ценную информацию: например, они могут сообщить, что «с 99%-ной уверенностью годовые убытки не превысят определенную сумму (VaR), но в худшем 1% случаев убытки могут быть чрезвычайно высокими». Вычисляя сводную статистику — средний убыток, медианный убыток, VaR на различных уровнях уверенности и ожидаемый дефицит — банк может решить, какой объем буфера капитала необходим для покрытия экстремальных событий. Результаты моделирования также облегчают анализ хвостового риска: менеджеры могут исследовать сценарии, которые приводят к самым большим потерям (например, одновременные дефолты крупных заемщиков или экстремальные рыночные обвалы) и разрабатывать планы действий в чрезвычайных ситуациях или стратегии

хеджирования. В целом, имитационное моделирование является незаменимым дополнением к аналитическим моделям, особенно когда факторы риска взаимодействуют сложным образом или когда закрытые решения не поддаются обработке. Как отмечено в обзоре Investopedia, имитации Монте-Карло «помогают объяснить влияние риска и неопределенности в моделях прогнозирования и предсказания», что делает их важнейшим инструментом в арсенале современного риск-менеджера.

1.3 Регуляторные аспекты в управлении банковскими рисками

Регуляторные аспекты в управлении банковскими рисками включают в себя требования надзорного органа - Центрального банка Российской Федерации, а также рекомендации Базельского комитета по банковскому регулированию и надзору. Среди основных видов и направлений регуляторных процедур, которые помогают управлять банковскими рисками, можно выделить:

1. закрепление порядка предоставления услуг и проведения операций во внутрибанковских нормативных документах;
2. недопустимость проведения операций, по которым отсутствуют утверждённые методики оценки и процедуры управления рисками;
3. использование системы лимитов, ограничивающих размеры портфелей и позиций, а также полномочия по принятию решений;
4. рейтинговые оценки клиентов, контрагентов, качества обеспечения по обязательствам заёмщиков;
5. регулярный мониторинг уровня рисков, принимаемых банком;
6. регулярное проведение стресс-тестирования.

Кроме того, регулятор определяет границы риска в банковской системе, в качестве такого регулятора выступает Банк России.

Закрепление порядка предоставления банковских услуг (продуктов) и проведения операций во внутрибанковских нормативных документах является одним из методов управления банковскими рисками. Такой подход позволяет:

1. минимизировать риски, связанные с несоблюдением должностными лицами соответствующих лимитов и превышением предоставленных им полномочий.

2. не допускать проведения банковских операций, приводящих к возникновению рисков, по которым отсутствуют утверждённые методики оценки и процедуры управления.
3. не принимать положительное решение о проведении банковской операции, предоставлении клиентам услуг (продуктов) без соблюдения предусмотренных внутрибанковскими документами надлежащих процедур.

Недопустимость проведения операций, по которым отсутствуют утверждённые методики оценки и процедуры управления рисками, является мерой по управлению банковскими рисками. Согласно Положению Банка России от 08.04.2020 №716-П «О требованиях к системе управления операционным риском в кредитной организации и банковской группе», кредитная организация должна утверждать процедуры управления операционным риском. Если в организации отсутствуют специализированные подразделения, процедуры управления отдельными видами операционного риска выполняет служба управления рисками. Также, согласно пункту 2.7 Указания Банка России №3624-У, кредитной организации необходимо обеспечить распределение между структурными подразделениями функций, связанных с принятием и управлением рисками, таким образом, чтобы осуществление операций (сделок), связанных с принятием рисков, и управление рисками не являлись функциями одного подразделения. Использование системы лимитов, ограничивающих размеры портфелей и позиций, а также полномочия по принятию решений, является одним из методов управления банковскими рисками. Основная задача системы установления лимитов — обеспечение формирования структуры активов и пассивов банка, адекватной характеру и масштабам его бизнеса.

Для минимизации кредитного риска на рынке межбанковского кредитования и рынке ценных бумаг риск-подразделением проводится анализ банков-контрагентов и эмитентов ценных бумаг с целью установления соответствующих лимитов.

Для регулирования рыночных рисков используются номинальные лимиты по видам финансовых инструментов, определяющие максимальный размер текущей позиции по ним на конец дня.

Для управления фондовым риском применяются оценка финансового состояния эмитента, установление лимитов на эмитентов ценных бумаг, лимитов на операции с ценными бумагами, установление срока вложений в финансовые инструменты.

Также разработка внутрибанковских лимитов и нормативов позволяет определить размер и количество рискованных операций, проводимых банком за один период.

Таким образом, регулятивные меры в области управления банковскими рисками направлены на увеличение размера собственных средств (капитала), улучшение его качества и обеспечение достаточного уровня покрытия капиталом принимаемых кредитными организациями рисков. Также благодаря реализации регулятивных мер в области управления банковскими рисками уделяется повышенное внимание вопросам идентификации и оценки рисков, в том числе имеющих системную природу. При этом осуществляется совершенствование банковского регулирования и банковского надзора, направленного на повышение качества банковского капитала и активов, ограничение уровня рисков, включая степень их концентрации, повышение транспарентности кредитных организаций.

1.4 Роль эконометрики в банковском риск-менеджменте

Эконометрика, как применение статистических методов к экономическим данным, играет центральную роль в современном банковском риск-менеджменте. Она обеспечивает теоретическую и количественную основу для построения моделей, связывающих факторы риска с результатами, и для вывода выводов из данных в условиях неопределенности. В контексте банковского риска эконометрические модели помогают оценивать вероятности дефолта, прогнозировать убытки, оценивать влияние макроэкономических условий и проверять гипотезы о факторах риска. Эконометрические методы в сочетании с растущей доступностью данных (часто называемых «большими данными» в финансах) позволяют риск-менеджерам выходить за рамки эвристических или экспертных подходов к более объективным, основанным на доказательствах методам.

Эконометрические модели бинарного выбора (например, модели логит или пробит), которые оценивают вероятность дефолта на основе различных объясняющих переменных. Такие модели являются основой управления кредитным риском как для отдельных заемщиков, так и для портфелей. Например, при моделировании кредитного портфеля или стресс-тестировании банковской системы пробит-модель может связывать вероятность дефолта фирмы с макроэкономическими переменными (рост ВВП, процентные ставки) и переменными, специфичными для фирмы (коэффициент левериджа, прибыльность). Оценивая коэффициенты, можно измерить чувствительность риска дефолта к каждому фактору. Эти модели часто дают совокупную меру риска — например, вероятность

дефолта банка или между банками при определенных условиях — которую можно использовать для оценки системной стабильности или влияния изменений политики. Банки часто оцениваются внешними агентствами (Moody's, S&P и т. д.) и эконометрические модели могут воспроизводить или анализировать эти рейтинги.

Академические исследования и специалисты по рискам создают модели прогнозирования рейтингов, которые берут финансовые коэффициенты банков (достаточность капитала, показатели качества активов, прибыль, коэффициенты ликвидности) и, возможно, макропеременные, чтобы предсказать долгосрочный кредитный рейтинг банка. Эти модели могут помочь определить, какие факторы оказывают наибольшее влияние на присвоение рейтингов (например, возможно, коэффициент капитала 1-го уровня, коэффициент неработающих кредитов и рентабельность активов являются значимыми предикторами), а также служат системами раннего оповещения, если прогнозируемый рейтинг банка начинает ухудшаться относительно фактического. Кроме того, сравнивая рейтинги разных агентств и то, что предполагает эконометрическая модель, аналитики могут обнаружить потенциальные несоответствия или предубеждения.

Эконометрический анализ используется для изучения того, как банки устанавливают процентные ставки по кредитам и депозитам, и как они связаны с принятием риска. Например, банки, которые занимаются более рискованным кредитованием, могут взимать более высокие процентные ставки, но эконометрические модели (например, модели одновременных уравнений или панельные регрессии) могут помочь распутать причину и следствие. Определенное направление исследований моделирует канал банковского кредитования денежно-кредитной политики: то есть, как изменения ставок центрального банка влияют на процентные ставки и объемы банковских кредитов, которые, в свою очередь, влияют на риск (например, замещение активов или достижение доходности). Некоторые эконометрические модели отмечают банки, которые, учитывая среду процентных ставок, похоже, берут на себя чрезмерный риск — возможно, выплачивая необычно высокие ставки по депозитам (что может указывать на отчаянную потребность в ликвидности) или взимая необычно высокие ставки по кредитам (что может указывать на кредитование очень рискованных заемщиков). Модели временных рядов (например, модели коррекции ошибок) также могут использоваться банками для прогнозирования подверженности риску процентной ставки, моделирования кривой доходности и того, как ее сдвиги влияют на чистый процентный доход и стоимость портфеля.

Эконометрика также используется для оценки эффективности работы банка (часто с помощью модели стохастической границы или анализа охвата данных, который, хотя и является более линейным программированием, чем эконометрика, может быть интегрирован с эконометрическим анализом).

Другое применение — использование эконометрических моделей для выявления факторов, влияющих на прибыльность банка (ROE, ROA) и риск (волатильность доходности, вероятность дефолта) — по сути, связывая риск и доходность. Регрессируя показатели прибыльности по прокси риска и контрольным переменным, аналитики могут понять компромиссы и выявить банки-выбросы (например, банк, получающий высокую прибыль с учетом своей капитализации, может быть либо очень эффективным, либо принимать скрытые риски, что является сигналом для более глубокого анализа).

Другая область — это обработка пропущенных данных и структурных разрывов, которые классическая эконометрика обрабатывает с помощью таких методов, как максимальное правдоподобие или множественное вменение, а машинное обучение иногда обрабатывает алгоритмическими средствами (деревья по своей сути могут обрабатывать пропущенные значения, разделяя их на «пропущенные» категории, например). Объединяя их, риск-менеджеры могут использовать предсказательную силу машинного обучения, одновременно используя эконометрические методы для проверки надежности и стабильности этих моделей с течением времени.

Эконометрический анализ также лежит в основе проверки модели. Когда модель риска банка построена (будь то регрессия или модель машинного обучения), команды проверки часто используют эконометрические подходы для проверки статистической надежности модели: они проверяют такие вещи, как стабильность вне выборки, отсутствие смещения пропущенных переменных, правильную калибровку (с помощью тестов Хосмера-Лемешоу или аналогичных для моделей PD) и дискриминационную способность (используя Джини или AUC, которые связаны с мерами ранговой корреляции). Они также могут переоценивать модели с помощью альтернативных методов, чтобы увидеть, сохраняются ли результаты (анализ чувствительности), что является эконометрическим упражнением по надежности.

Глава 2: Машинное обучение и многоагентные модели в управлении банковскими рисками

2.1 Обзор методов машинного обучения, применяемых в банковской сфере

Машинное обучение (ML) стало краеугольным камнем аналитических методов в банковском секторе, поддерживая приложения от оценки кредитного риска и обнаружения мошенничества до сегментации клиентов и алгоритмической торговли. Традиционно банки полагались на статистические модели (например, логистическую регрессию), но в последние годы они все чаще применяют более сложные методы ML для повышения точности прогнозирования. Ключевые алгоритмы контролируемого обучения, такие как деревья решений, машины опорных векторов (SVM), нейронные сети и ансамблевые методы (например, случайные леса и градиентный бустинг), теперь играют важную роль в финансовом моделировании. Эти алгоритмы могут улавливать нелинейные закономерности и взаимодействия в больших наборах данных, часто превосходя более простые модели в таких задачах, как прогнозирование дефолта или оповещения о борьбе с отмыванием денег. В то же время банковская сфера предъявляет строгие требования к интерпретируемости и надежности моделей, что влияет на выбор и разработку методов ML.

Деревья решений — это интуитивно понятные, но мощные модели классификации и регрессии, которые разделяют данные с помощью последовательности бинарных разделений. Дерево состоит из узлов (представляющих решения по признаку) и узлов листьев (конечные результаты или прогнозы). Начиная с корневого узла, дерево рекурсивно разделяет набор данных на основе пороговых значений признаков, которые максимизируют однородность целевой переменной в полученных подмножествах. Это жадное разделение продолжается до тех пор, пока не будут выполнены критерии остановки (например, максимальная глубина дерева или минимальный размер листа), и на каждом листе делается прогноз большинством голосов (классификация) или средним значением (регрессия) обучающих выборок в этом листе. Деревья решений имитируют процессы принятия решений человеком с помощью серии правил «если-то», что делает их легко интерпретируемыми. Они могут фиксировать нелинейные отношения и

взаимодействия переменных, не требуя нормализации данных или предположений о линейности. Классический алгоритм построения деревьев решений — CART (деревья классификации и регрессии), представленный Брейманом, в котором для выбора оптимальных разбиений используются двоичные разбиения и такие меры, как коэффициент Джини или прирост информации.

Чтобы решить, по какому признаку и порогу разделять в узле, алгоритмы дерева определяют меру примеси. Одной из распространенных метрик для классификации является примесь Джини, определяемая как:

$$I_G(p) = 1 - \sum_{k=1}^K p_k^2,$$

где p_k — доля обучающих выборок класса k в этом узле. I_G равен 0, когда все выборки принадлежат одному классу (чистые), и больше 0, когда смесь классов более равномерна. Другой популярной метрикой является энтропия:

$$H = - \sum_k p_k \cdot \log_2 p_k,$$

а соответствующий прирост информации от разделения — это уменьшение энтропии:

$$\begin{aligned} \text{InfoGain}(T, \text{деление}) &= H(\text{родительский узел}) \\ &- \sum_{\text{сестринский узел } j} \frac{N_j}{N_{\text{родитель}}} \cdot H(\text{сестринский узел } j), \end{aligned}$$

где N_j — это число образцов в дочернем узле j . Алгоритм выберет разбиение, которое максимизирует прирост информации (или, что эквивалентно, минимизирует примесь в дочерних узлах). Рекурсивно применяя такие разбиения, дерево разбивает пространство признаков на непересекающиеся области.

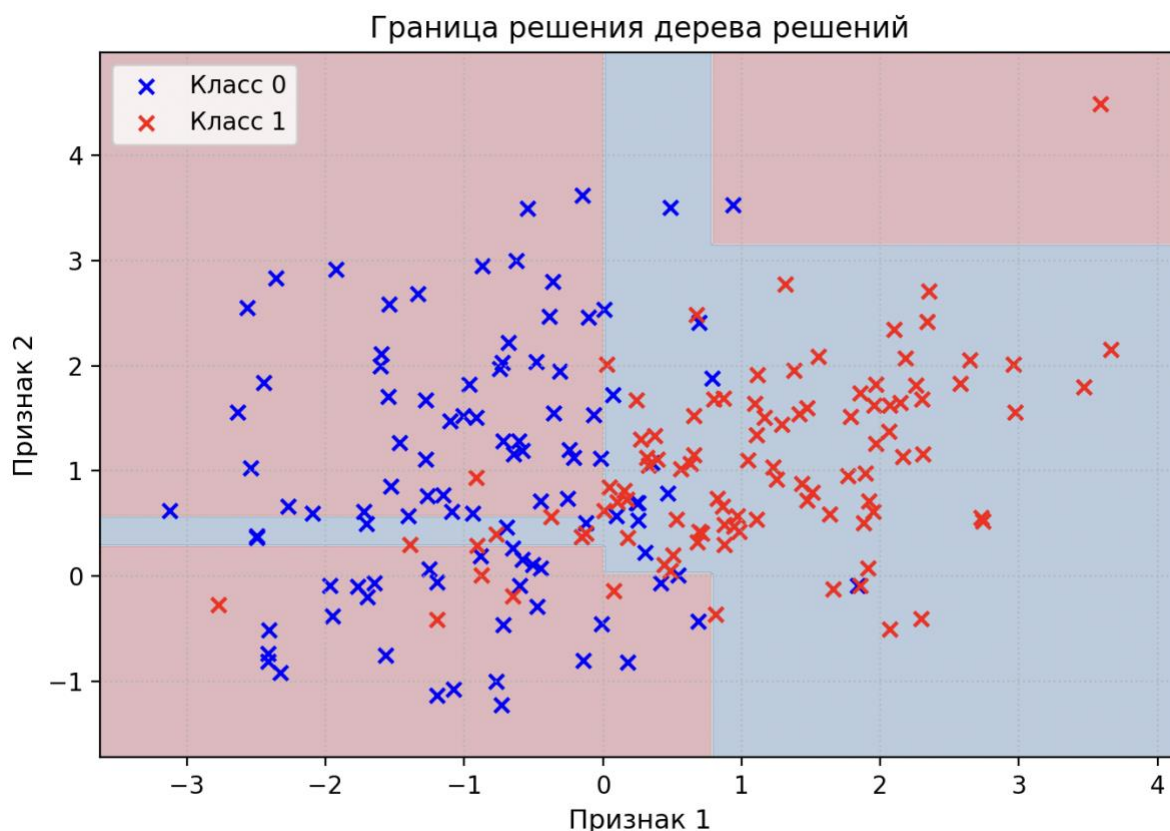


Рисунок 9. Пример границы решения модели дерева решений на двумерном наборе данных.

Дерево разбивает пространство признаков на прямоугольные области (закрашенные синим для одного класса и красным для другого), устанавливая пороги для Признака 1 и Признака 2. Иерархическая логика *if-then* дерева решений создает кусочно-постоянные области прогнозирования, как показано резкими границами между красными и синими областями.

Деревья решений имеют несколько преимуществ, которые сделали их популярными в финансах. Это модели «белого ящика» — обоснование прогноза можно отследить по пути в дереве (полезно для объяснения кредитных решений или одобрений займов регулирующим органам и клиентам). Они изначально обрабатывают смешанные типы данных и не чувствительны к выбросам или масштабированию. В банковском деле деревья решений использовались для кредитного скоринга и классификации рисков, давая простые правила принятия решений (например, «ЕСЛИ доход < 50 тыс. долл. И кредитная история = плохая, ТО риск = высокий»). Они также могут улавливать эффекты взаимодействия, такие как объединение нескольких атрибутов клиента для определения риска. Однако одно дерево решений может быть нестабильным и склонным к переобучению — небольшие изменения в данных могут изменить структуру дерева, а слишком глубокое дерево будет соответствовать шуму. Методы обрезки (уменьшение

глубины дерева или требование минимального размера узла) используются для улучшения обобщения. Несмотря на эти ограничения, деревья решений часто служат строительными блоками для более мощных ансамблевых методов, а их интерпретируемость хорошо согласуется с нормативным акцентом в банковской сфере на прозрачных моделях.

Опорные векторы — это класс мощных моделей контролируемого обучения, используемых для классификации (и регрессии), которые находят оптимальную разделяющую гиперплоскость между классами. SVM возникли из статистической теории обучения и особенно хорошо обоснованы в математике — они стремятся максимизировать разницу между различными классами.

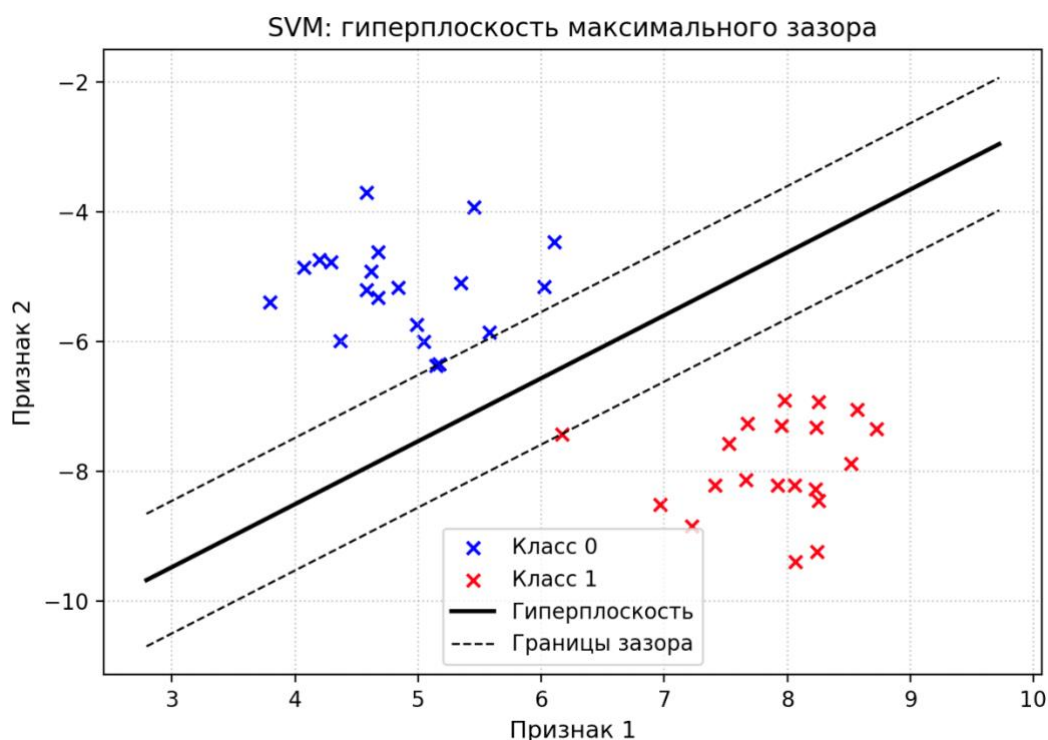


Рисунок 10. SVM, находящая линейный разделитель с максимальным запасом. Граница принятия решения (сплошная линия) расположена посередине между ближайшими точками двух классов (синие крестики против красных крестиков). Запас — это расстояние между границей и этими ближайшими точками, показанное параллельными пунктирными линиями. Точки на запасе — это опорные векторы, которые определяют оптимальную гиперплоскость (Cortes & Vapnik, 1995).

Главной силой SVM является трюк с ядром, который позволяет выполнять нелинейную классификацию путем неявного сопоставления данных с многомерным пространством

признаков. Используя функции ядра (например, гауссовскую RBF, полиномиальную), SVM может найти линейный разделитель в расширенном пространстве признаков без явного вычисления координат в этом пространстве, эффективно создавая нелинейные границы решений в исходном пространстве. Это делает SVM очень гибкими: например, при обнаружении мошенничества или кредитного риска SVM с ядром RBF может захватывать сложные нелинейные закономерности в данных о транзакциях клиентов.

Однако SVM имеют некоторые недостатки. Они могут быть вычислительно интенсивными для больших наборов данных (время обучения растет сверхлинейно с числом образцов в худшем случае), и модель не так легко интерпретируется, как дерево решений или линейная регрессия — функция решения определяется опорными векторами в многомерном пространстве, что не так просто объяснить заинтересованным сторонам.

Нейронные сети (также известные как искусственные нейронные сети, ANN) — это класс моделей, вдохновленных нейронной архитектурой человеческого мозга. Они состоят из слоев взаимосвязанных нейронов (или узлов), которые преобразуют входные данные посредством взвешенных связей и нелинейных функций активации. Простейшей архитектурой является нейронная сеть прямого распространения (также называемая многослойным персептроном, MLP), где данные передаются из входного слоя через один или несколько скрытых слоев в выходной слой. Каждый нейрон j в слое вычисляет взвешенную сумму выходов из предыдущего слоя,

$$z_j = \sum_i w_{ij} \cdot x_i + b_j,$$

а затем применяет нелинейную активацию $\sigma(\cdot)$ (такую как сигмоид $\sigma(z) = \frac{1}{1+e^{-z}}$) для получения $a_j = \sigma(z_j)$.

Например, сеть с одним скрытым слоем с входом $x \in R^d$, скрытым слоем нейронов H и выходом \hat{y} может быть выражена как:

$$\hat{y} = \phi\left(\sum_{j=1}^H v_j \cdot \sigma\left(\sum_{i=1}^d w_{ij} \cdot x_i + b_j\right) + c\right),$$

Где w_{ij} — веса вход-скрытый, — v_j веса скрытый-выход, b_j и c — смещения, $\sigma(\cdot)$ — активация, а $\phi(\cdot)$ — выходная активация. Сила нейронных сетей заключается в их

способности изучать богатые, нелинейные представления данных; фактически, сеть даже с одним скрытым слоем может аппроксимировать любую непрерывную функцию на компактной области с произвольной точностью при наличии достаточного количества нейронов (универсальная теорема аппроксимации).

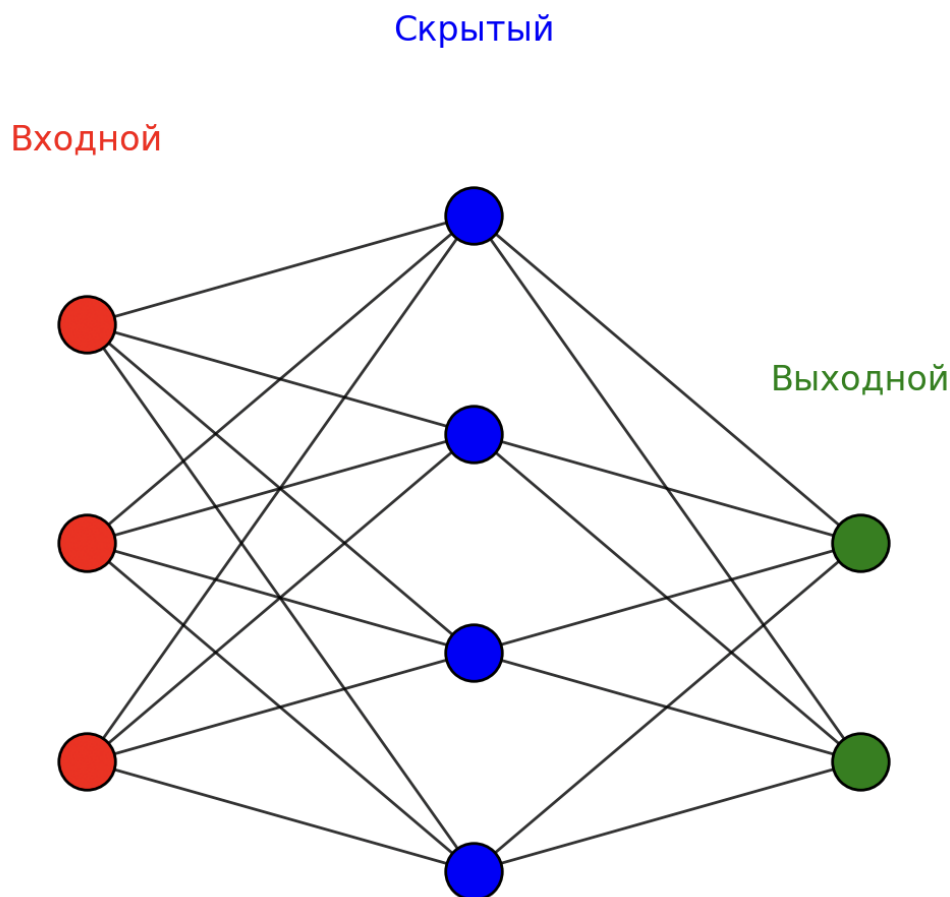


Рисунок 11: Простая архитектура нейронной сети прямого распространения. Она содержит входной слой (слева, красные узлы), один скрытый слой (посередине, синие узлы) с полными связями из входного слоя и выходной слой (справа, зеленые узлы). Каждая стрелка представляет собой обученный вес, соединяющий нейроны между слоями. Такие сети обучаются путем корректировки весов для минимизации ошибки между прогнозами и истинными выходами.

Нейронные сети «учатся» на данных с помощью процесса, называемого обучением, обычно с использованием обратного распространения в сочетании с оптимизацией на основе градиента. Во время обучения веса сети итеративно корректируются для минимизации выбранной функции потерь (например, среднеквадратической ошибки для регрессии или перекрестной энтропии для классификации). Обратное распространение эффективно вычисляет градиент потерь относительно всех весов, распространяя ошибку

обратно через слои сети. Повторение этого на протяжении многих эпох позволяет сети корректировать веса для соответствия сложным закономерностям в данных. Современные нейронные сети со множеством скрытых слоев (глубокое обучение) достигли выдающихся успехов в таких областях, как компьютерное зрение и обработка естественного языка, благодаря улучшениям в алгоритмах, методах регуляризации и аппаратном ускорении.

В банковском секторе нейронные сети имеют долгую историю использования в таких задачах, как кредитный скоринг, где ранние исследования показали, что ИНС могут конкурировать или превосходить логистическую регрессию в прогнозировании дефолтов по кредитам. Они также использовались для обнаружения мошенничества, где сеть может изучать тонкие, нелинейные комбинации признаков транзакций, которые указывают на мошенническое поведение. Совсем недавно глубокие нейронные сети были применены к крупномасштабным финансовым данным — например, рекуррентные нейронные сети и модели LSTM для моделирования последовательностей в потоках транзакций по кредитным картам или глубокие сети прямого распространения для прогнозирования просрочки по ипотеке.

Несмотря на точность, ключевой проблемой нейронных сетей в банковском деле является отсутствие интерпретируемости. Их часто называют «черными ящиками», поскольку изученные веса и активации не поддаются простому объяснению — проблема, когда регулирующие органы требуют четких причин для кредитных решений или оценок рисков. Это подстегнуло исследования в области объяснимого ИИ; например, использование значений SHAP или LIME для интерпретации прогнозов нейронных сетей для отдельных заемщиков. Еще одним практическим соображением является то, что нейронным сетям обычно требуются большие объемы данных и обширная настройка гиперпараметров (слоев, нейронов, скоростей обучения и т. д.) для хорошей работы, что может быть ресурсоемким. Однако, когда данных много и точность имеет первостепенное значение, нейронные сети являются убедительным выбором и продолжают набирать обороты в сфере финансов.

Методы ансамбля объединяют несколько базовых моделей для создания более надежной и точной. Два наиболее распространенных метода ансамбля в банковском деле — это методы бэггинга, такие как случайные леса, и методы бустинга, такие как градиентный бустинг. Оба подхода основаны на базовых моделях дерева решений, используя их гибкость и одновременно смягчая их тенденцию к переобучению. Модели ансамбля

демонстрируют высочайшую производительность в соревнованиях с табличными данными и в реальных задачах, включая многие финансовые приложения.

Случайный лес по сути является набором (ансамблем) деревьев решений, где каждое дерево обучается на случайном подмножестве данных и признаков. Деревья «голосуют» за окончательное предсказание. Для классификации лес предсказывает класс, который является наиболее частым среди выходных данных деревьев (большинство голосов), а для регрессии он предсказывает среднее значение предсказаний деревьев.

Два ключевых шага рандомизации придают случайным лесам их силу: каждое дерево обучается на выборке (выборка с заменой, поэтому каждое дерево видит немного другой набор данных того же размера), и при каждом разделении дерева рассматривается случайное подмножество признаков (вместо всех признаков). Этот случайный выбор признаков («бэггинг признаков» Бреймана) декоррелирует деревья — поскольку не все деревья будут разделяться на одних и тех же доминирующих признаках, их ошибки более независимы. Агрегация многих декоррелированных деревьев приводит к сильному снижению дисперсии прогноза ансамбля, сохраняя при этом низкое смещение. Случайные леса часто достигают превосходной точности из коробки и требуют небольшой настройки параметров (помимо количества деревьев). Они также дают полезные оценки важности признаков, измеряя, насколько каждый признак способствует снижению примеси во всем ансамбле.

В банковском секторе случайные леса широко применяются для моделирования кредитного риска, обнаружения мошенничества и маркетинговой аналитики. Например, в кредитном скоринге случайные леса превосходили логистическую регрессию и отдельные деревья решений, фиксируя нелинейные взаимодействия между атрибутами заемщика. Они также относительно хорошо справляются с несбалансированными наборами данных и пропущенными значениями — важные черты для реальных банковских данных. Исследования показали, что ансамбли на основе деревьев, такие как случайные леса и усиленные деревья, дают превосходную производительность в контрольных показателях кредитного скоринга. Более того, хотя случайный лес менее интерпретируем, чем отдельное дерево решений, графики частичной зависимости и рейтинги важности признаков все еще могут дать представление о том, какие факторы управляют прогнозами, предлагая компромисс между точностью и объяснимостью.

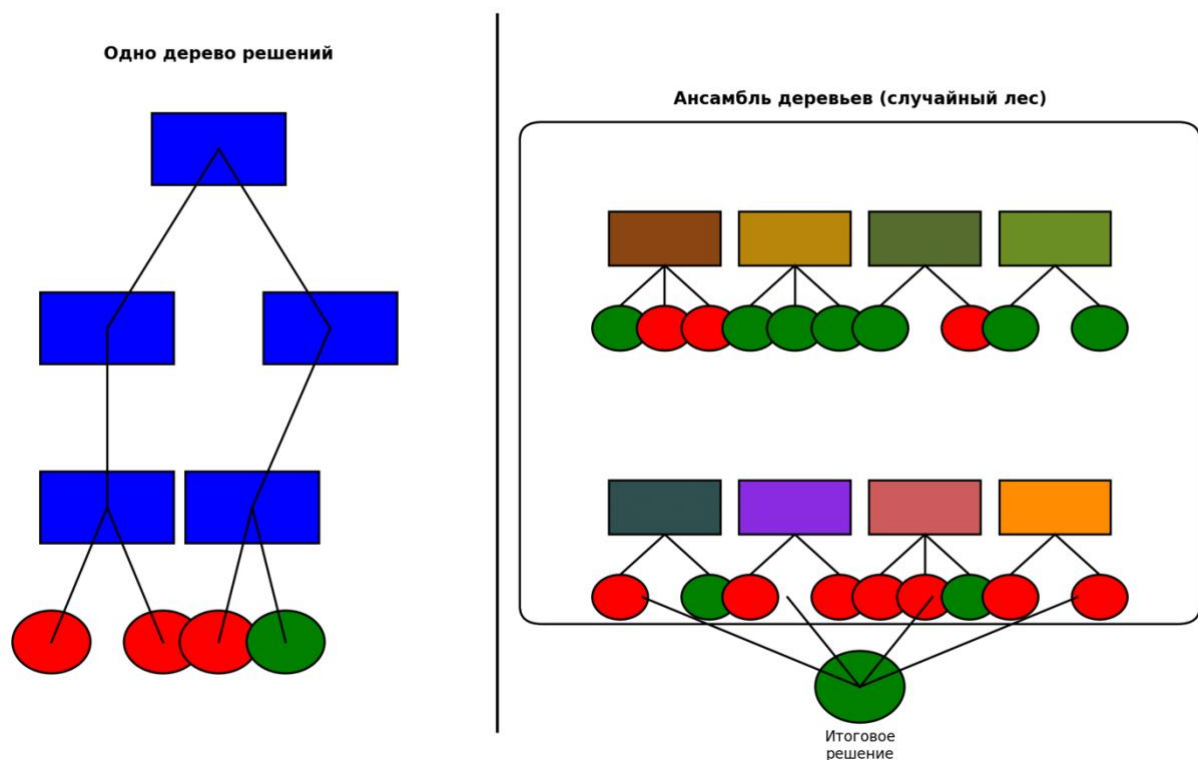


Рисунок 12. Иллюстрация ансамблевого обучения в случайном лесу. Слева - одно дерево решений может допускать ошибки (красные или зеленые узлы указывают на неправильную классификацию). Справа: ансамбль из множества разнообразных деревьев (каждое с различными разделениями из-за случайной выборки) голосует за окончательный результат. Совокупный прогноз (большинство голосов за классификацию) точнее, чем любое отдельное дерево, поскольку ошибки деревьев компенсируются. Случайные леса используют этот принцип для повышения стабильности и точности по сравнению с отдельными деревьями решений.

Градиентный бустинг использует другой подход - вместо того, чтобы обучать много больших деревьев независимо, оно последовательно строит ансамбль, где каждое новое дерево исправляет ошибки текущего ансамбля. Градиентное усиление является известным методом усиления, который рассматривает обучение ансамбля как аддитивную модель и использует градиентный спуск в функциональном пространстве для минимизации функции потерь.

В банковском деле градиентный бустинг известен своей высокой эффективностью при решении различных задач. Он стал популярным методом в аналитике кредитного риска, часто обеспечивая наилучшую точность в прогнозировании дефолта по кредиту или

оттока клиентов. Например, усиленные деревья могут улавливать тонкие нелинейные эффекты: небольшое увеличение отношения долга к доходу может резко повысить риск дефолта только после определенного порогового значения, и алгоритмы усиления могут моделировать такие пороговые эффекты путем поэтапной подгонки. Эмпирические исследования подтверждают, что методы усиления (например, AdaBoost, градиентный бустинг) обычно входят в число лучших алгоритмов для кредитного скоринга. Они также использовались для борьбы с отмыванием денег (выявление подозрительных транзакций) и моделей рыночного риска, где решающее значение имеет способность подгонять сложные закономерности в данных. Одним из недостатков является то, что усиленные модели, такие как нейронные сети, сложны и не поддаются интерпретации; однако специалисты-практики смягчают это с помощью таких инструментов, как значения SHAP, для интерпретации влияния признаков даже в этих ансамблях.

С точки зрения регулирования ансамблевые модели в банковском деле вызывают проблемы, поскольку их сложность может затруднить проверку и объяснение. Тем не менее, их превосходная точность часто является убедительной для бизнес-кейсов, поэтому банки используют их с осторожностью: обширная проверка, мониторинг предвзятости и иногда использование отклонений вывода или аппроксимации оценочной карты для перевода модели в более интерпретируемые термины для соответствия. Исследования в области «объяснимого ML» особенно активны для этих моделей, направленных на поддержание высокой производительности при удовлетворении требований прозрачности финансовых правил.

2.2 Применение методов машинного обучения для моделирования кредитных рисков

Методы машинного обучения (*ML*) все чаще применяются в моделировании кредитного риска для повышения точности прогнозирования, скорости принятия решений. Например, Сбербанк применил ансамблевые модели *ML* в 2018 году. Эти модели используют большие данные (включая записи транзакций клиентов) для оценки вероятности дефолта (*PD*) с помощью оценочной карты на основе логистической регрессии, улучшенной с помощью градиентно-усиленных деревьев решений (метод AutoWOE, объединяющий логистическую регрессию с усиленными деревьями). Например, *PD* для заемщика *i* можно смоделировать с помощью логистической функции:

$$PD_i = \frac{1}{1 + \exp(-(\beta_0 + \sum_i \beta_i x_i))}$$

где признаки x_{ij} включают недавний оборот по счету, платежное поведение и т. д.

Нелинейные алгоритмы машинного обучения (например, XGBoost) фиксируют сложные взаимодействия между этими признаками, дополняя базовую логистическую модель.

Используя такие модели, автоматизированная платформа андеррайтинга Сбербанка может одобрять кредиты до 2 млрд руб. за 7 минут с минимальным ручным вводом. Это сократило время принятия решения о выдаче кредита с нескольких дней до нескольких минут, что дало Сбербанку конкурентное преимущество. К концу 2018 года система на основе машинного обучения была развернута в регионах, что помогло увеличить корпоративный кредитный портфель более чем на 12% в годовом исчислении при сохранении строгих стандартов риска. Быстрый скоринг на основе машинного обучения увеличил показатели одобрения и суммы кредитов для кредитоспособных заемщиков без повышения уровней дефолта, поскольку модели оценивают вероятность дефолта каждой заявки в режиме реального времени и одобряют только в том случае, если прогнозируемый риск ниже приемлемых пороговых значений. Отраслевые исследования подтверждают, что такие ансамблевые методы превосходят традиционные модели — например, экстремальное градиентное усиление улучшило индекс Джини (точность ранга риска) примерно на 5 процентных пунктов по сравнению с традиционной логистической регрессией.

UniCredit, крупный европейский банк, объединился с Prometeia для разработки модели PD на основе ML для своего портфеля МСП (малые и средние предприятия). Эта инициатива (около 2020 г.) стала пионером в использовании гранулярных данных о транзакциях и текстовой информации в моделировании кредитного риска. Эти обогащенные признаки (шаблоны транзакций, волатильность денежных потоков, сетевые индикаторы и т. д.) были включены в мощные нелинейные модели (например, глубокое обучение, градиентный бустинг), что дало оценку «транзакционного» риска для каждой фирмы. Окончательная модель (классификатор на основе XGBoost) сопоставляет поведение транзакций с высокой размерностью PD, в то время как объяснимые методы ИИ обеспечивают интерпретируемость. UniCredit интегрировала эту оценку ML в свой кредитный рабочий процесс в качестве наложения на традиционную систему финансового рейтинга. Теперь у каждого МСП есть обновленный с помощью ИИ PD наряду с обычным рейтингом, и специалисты по рискам используют его как при андеррайтинге, так и при мониторинге. Во время COVID-19, когда многие заемщики находились под мораторием (маскируя традиционные предупреждающие знаки), модель ML оказалась бесценной, обнаружив тонкие признаки неблагополучия (например, резкое падение доходов,

задержка платежей) до любых пропущенных платежей. Это раннее предупреждение позволило оказать упреждающую поддержку или реструктуризацию, предотвращая дефолты. На практике UniCredit увидела улучшение дифференциации рисков — клиенты, отмеченные моделью ML как высокорисковые, имели значительно более высокие последующие ставки дефолта, чем те, которые были отмечены как низкорисковые, — и смогла безопасно предоставить больше кредитов малорисковым МСП с низким уровнем риска, которые ранее были недооценены. Такой рост точности и оперативности помог сократить новые неработающие кредиты и оптимизировать профиль риска и вознаграждения портфеля МСП. На рисунке 13 показано, как включение признаков, полученных из транзакций, сместило калибровочную кривую PD ближе к идеальной линии 45°, улучшив соответствие между прогнозируемыми вероятностями и фактическими показателями дефолта.

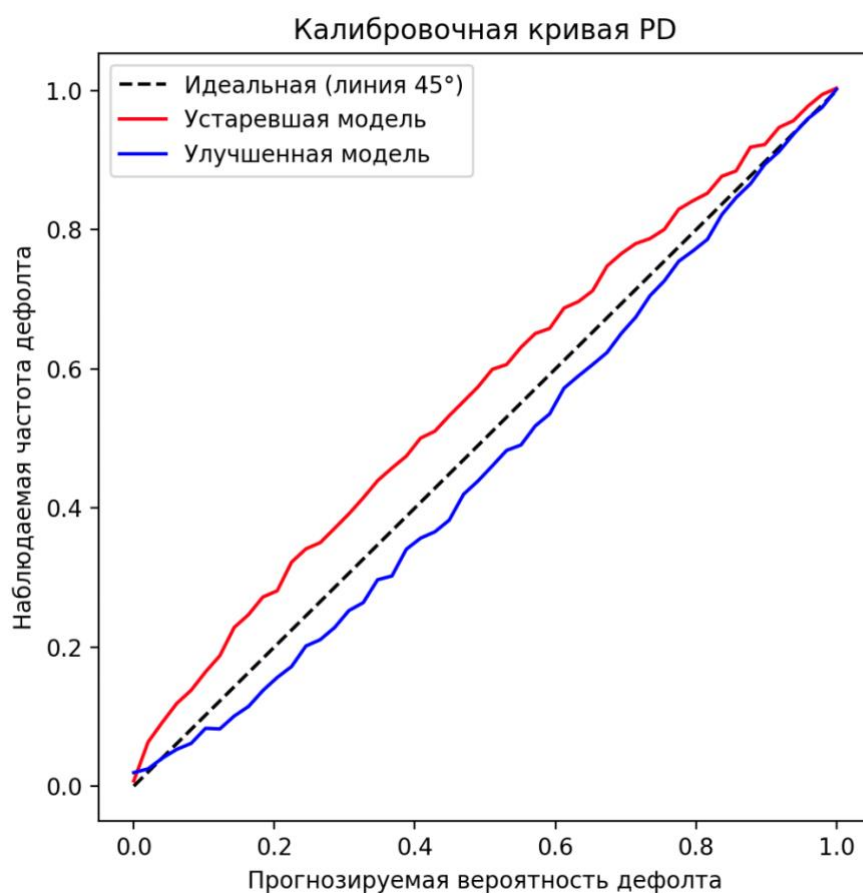


Рисунок 13. Улучшенная калибровка PD с детализированными данными о транзакциях. Кривая калибровки модели с улучшенным ML (синяя) лежит ближе к идеальной линии 45° (пунктирная), чем у устаревшей модели (красная), что указывает на то, что

прогнозируемые вероятности дефолта модели ML лучше соответствуют наблюдаемым частотам дефолта среди заемщиков.

BBVA реализовала передовой набор моделей ML для усиления управления кредитными рисками, сосредоточившись на раннем обнаружении дефолтов и оптимизации взысканий. В 2024 году AI Factory BBVA разработала набор превентивных моделей ML, прогнозирующих различные стадии просрочки по кредитам: например, вероятность того, что текущий кредит станет просроченным, что просроченный счет будет вылечен или что дефолт сохранится. Разработка началась с ~1800 признаков (демографические данные клиентов, транзакции, кредитная история, макроэкономические данные), сокращенных с помощью отбора признаков для устранения избыточности. Команда сравнила простую логистическую регрессию (для интерпретируемости) с более сложными моделями (случайные леса, XGBoost, нейронные сети), обнаружив, что последние значительно повышают прогностическую эффективность. Используя автоматическую настройку гиперпараметров (Optuna) для максимизации AUC и избежания переобучения, они выбрали оптимальные алгоритмы для каждой подмодели. Окончательные модели показали высокую дискриминацию (сильные коэффициенты Джини) и были проверены на несвоевременных данных для надежности. BBVA интегрировала эти модели ML в свою систему раннего предупреждения (EWS) и рабочий процесс взыскания. Модели регулярно работают с данными о кредитах и генерируют оповещения о рисках для упреждающих действий. Например, система помечает счета с высокой вероятностью просрочки, чтобы BBVA могла связаться с ними до пропущенного платежа (например, предлагая помощь или изменяя условия). Другие модели направляют взыскание, различая случаи, которые, скорее всего, будут излечены самостоятельно, от случаев с высоким риском долгосрочного дефолта, что позволяет эффективно расставлять приоритеты в усилиях по взысканию. Этот динамический подход, основанный на данных, вытеснил исключительную зависимость от статических правил, и весь конвейер ML автоматизирован от начала до конца для быстрого обновления по мере поступления новых данных. BBVA сообщает, что модели ML EWS существенно улучшили раннее обнаружение рисков и точность управления задолженностью. Внутренние тесты показали более высокие AUC на каждом этапе (например, значительно лучшие 3-месячные прогнозы дефолта с меньшим количеством ложных тревог). Следовательно, пилотные портфели показали более низкие показатели просрочки и дефолта благодаря более ранним целенаправленным вмешательствам. Примечательно, что объемы просрочки более 90 дней снизились, поскольку многие клиенты из группы риска получили индивидуальные

решения в окне просрочки от 1 до 30 дней, что предотвратило дальнейшее проскальзывание. Показатели взыскания по просроченным кредитам также выросли, поскольку модели определили, какие неплательщики могут возобновить платежи при соответствующей реструктуризации. Общим результатом стал более жесткий контроль кредитного риска (и, вероятно, сокращенные резервы на возможные потери по кредитам), что подтверждает инициативу BBVA как «яркий пример того, как ИИ трансформирует управление кредитными рисками». Рисунок 14 иллюстрирует повышение производительности: кривая ROC модели ML BBVA (синяя) изгибается ближе к верхнему левому углу, чем у устаревшей модели (серая), что указывает на более высокий показатель истинно положительных результатов для любого заданного показателя ложноположительных результатов (т. е. больший AUC).

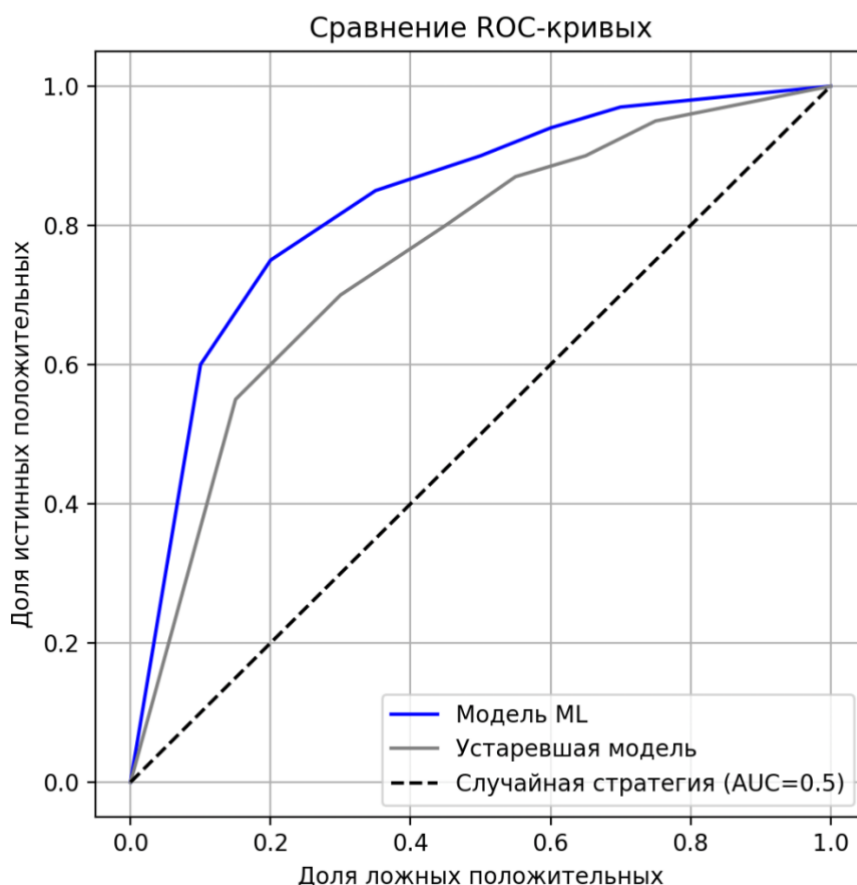


Рисунок 14. Концептуальные кривые ROC для модели раннего оповещения BBVA по сравнению с устаревшей моделью. Модель на основе машинного обучения (синяя) обнаруживает большую долю будущих дефолтеров на каждом уровне ложной тревоги, чем традиционная модель (серая), как показано ее кривой ROC, смещающейся ближе к

верхнему левому углу (более высокий AUC). Диагональная линия (45°) представляет собой случайное угадывание.

HSBC внедрили ML для управления кредитными лимитами по кредитным картам и связанным с ними риском. В 2020 году, на фоне потрясений, вызванных COVID-19, группа аналитиков рисков HSBC и FICO внедрила решение по оптимизации кредитной линии на основе искусственного интеллекта. Стратегия заключалась в использовании прогнозов ML для адаптации увеличения кредитной линии (CLI) для отдельных клиентов, уравнивая прирост дохода с кредитным риском. HSBC разработал прогностические модели для оценки увеличения расходов каждого клиента при повышении его лимита (потенциальный прирост дохода) и вероятности дефолта, связанной с более высоким лимитом. Эти модели (например, логистические или древовидные классификаторы, обученные на исторических результатах CLI) предоставляли управляемые данными входные данные вместо универсальных правил. Затем математический механизм оптимизации (платформа FICO) оценил различные решения CLI и определил оптимальный для портфеля набор — максимизируя общие расходы по карте, сохраняя общий риск в пределах аппетита банка. HSBC интегрировал это решение на основе ML в свое обычное управление счетами: вместо ручных или общих политик CLI система периодически запускает оптимизатор, рекомендуя, какие счета должны получить увеличение лимита на основе прогнозируемого профиля риска-вознаграждения каждого клиента. Строгие ограничения гарантировали соблюдение лимитов риска и справедливости (например, отсутствие чрезмерных увеличений, соблюдение нормативных требований). HSBC сообщил о 15%-ном росте средних ежемесячных расходов по карте среди клиентов, получивших CLI, по сравнению с предыдущей стратегией, что привело к более высокому процентному и комиссионному доходу. Что особенно важно, это было достигнуто без какого-либо роста безнадежной задолженности — показатели просрочки и убытки остались стабильными (или даже улучшились) по сравнению с контрольными группами. Благодаря разумному выбору клиентов с низким уровнем риска для CLI система ML избежала чрезмерного кредитования рискованных заемщиков, поэтому ключевые показатели риска (например, коэффициент неработающих кредитов) не ухудшились. В то же время улучшились показатели вовлеченности клиентов: активное использование карт и доля кошелька увеличились, поскольку клиенты, получившие более высокие лимиты, больше использовали свои карты HSBC. Этот успех принес HSBC премию FICO Decisions Award 2024 за ИИ в кредитном риске. Как резюмирует FICO, оптимизация ИИ достигла «золотой середины» между риском и вознаграждением,

скорректировав кредитную линию каждого клиента с научной точностью. Рисунок 15 иллюстрирует влияние риска и вознаграждения: новая стратегия, основанная на ML (зеленая точка), достигла примерно на 15% более высоких расходов на счет без увеличения уровня убытков портфеля, превзойдя предыдущую стратегию (серая точка), которая имела более низкие расходы при том же уровне риска.

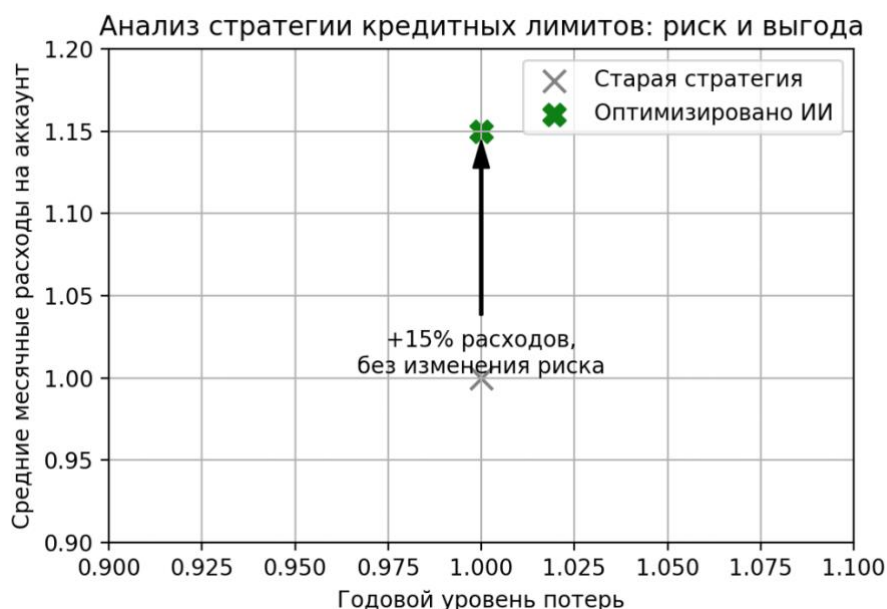


Рисунок 15. Результат риска и вознаграждения от оптимизации кредитного лимита HSBC. На этой наглядной диаграмме (ось x = годовой уровень убытков, ось y = ежемесячные расходы на счет) стратегия, оптимизированная с помощью ML (зеленая), обеспечивает более высокие расходы при том же уровне убытков, что и устаревшая стратегия (серая), тем самым улучшая позицию портфеля по риску и доходности.

Эти случаи демонстрируют, как методы машинного обучения могут значительно улучшить моделирование кредитного риска. Используя передовые алгоритмы (ансамблевые деревья, нейронные сети, оптимизация), такие банки, как Сбербанк, UniCredit, BBVA и HSBC, достигли более точных прогнозов *PD*, более раннего обнаружения ухудшения кредитоспособности и оптимизированных решений о кредитовании. Результаты включают более быстрое одобрение кредитов, сокращение дефолтов или просрочек и лучшие результаты соотношения риска и вознаграждения. Не менее важно, что эти модели ML были развернуты с вниманием к интерпретируемости и управлению, что гарантирует их эффективное дополнение к традиционному управлению рисками. Эти реальные реализации подчеркивают, что при разумном применении модели

кредитного риска на основе ML могут превзойти традиционные подходы и обеспечить ощутимые улучшения качества портфеля и прибыльности.

2.3 Введение в многоагентное моделирование банковских рисков

Многоагентные системы (MAS) являются вычислительными системами, в которых несколько автономных сущностей (агентов) взаимодействуют в определенной среде. Каждый агент действует на основе собственного набора правил или стратегий и может адаптироваться или принимать решения в ответ на действия других агентов и окружающую среду. Коллективное поведение системы возникает из множества взаимодействий, часто приводящих к сложной нелинейной динамике, которую трудно предсказать, анализируя любого отдельного агента в изоляции. В контексте финансового моделирования агентами могут быть отдельные лица или учреждения (например, потребители, банки, инвесторы), каждое из которых имеет свои собственные характеристики и процессы принятия решений.

В теоретическом плане моделирование MAS опирается на междисциплинарные основы, включая теорию игр, теорию сетей и науку о сложности. Агенты могут следовать эвристическим или ограниченно рациональным стратегиям, а не идеальной оптимизации, отражая реальное принятие решений в условиях неопределенности. Поскольку агенты взаимодействуют через сети (например, межбанковские сети кредитования или социальные сети заемщиков), связность играет решающую роль в распространении риска. Например, коллективный эффект от того, что многие банки управляют своими собственными рисками (например, отзывая кредиты или продавая активы), может непреднамеренно увеличить системный риск — явление, которое иногда называют парадоксом осмотрительности или сетевым внешним эффектом.

Другим важным теоретическим аспектом является то, как MAS может включать адаптивное поведение и петли обратной связи. Агенты могут быть спроектированы для обновления своих стратегий на основе опыта (обучения) или для адаптации к нормативным правилам. Эта адаптивность означает, что система может развиваться с течением времени, потенциально демонстрируя сдвиги режима или новые модели, если условия изменятся. Такие функции хорошо подходят для моделирования кредитного риска, который часто включает обратную связь (например, ужесточение условий кредитования после убытков, приводящее к изменениям в поведении заемщика). В результате многоагентные модели рассматриваются как способ зафиксировать динамику и

эндогенное накопление риска, которые традиционные статические модели могут упустить.

Традиционные модели кредитного риска в банковском деле, такие как кредитные скоринговые карты или модели портфеля, обычно фокусируются на количественной оценке риска при статических предположениях — например, оценивая вероятности дефолта и корреляции из исторических данных. Они часто рассматривают корреляции или заражение экзогенным образом или предполагают распространение шока сверху вниз. MAS, напротив, эндогенизируют распространение риска посредством явного моделирования взаимодействий. Вместо того чтобы предполагать заранее определенный параметр корреляции или заражения, модель на основе агентов может генерировать заражение в результате сетевого соединения и поведения агентов.

Одним из применений MAS является моделирование кредитного риска портфелей заемщиков, таких как держатели кредитных карт или корпоративные клиенты по кредитам. В таких моделях каждый заемщик может быть представлен как агент с определенными финансовыми характеристиками и правилами поведения (например, склонностью к погашению или дефолту в зависимости от экономических условий). Банки или кредитные специалисты также могут быть агентами, принимающими решения о кредитовании или управляющими кредитами с течением времени. Моделируя взаимодействия (заемщики, осуществляющие платежи или решения о дефолте, и банки, реагирующие путем корректировки кредитных лимитов или взысканий), можно оценить динамику кредитного риска портфеля при различных сценариях.

Шау Ли, Ху Чао и Лау Хуи разработал многоагентное моделирование портфеля кредитных карт банка для решения проблемы, связанной с тем, что банки не могут в полной мере отслеживать кредитоспособность каждого держателя карты. В их модели агенты-держатели карт классифицируются по типам (например, «склонные к овердрафту» и «не склонные к овердрафту») с различным поведением. Моделирование развивается в течение 36 месяцев, отслеживая, сколько агентов не выполняют обязательства при различных начальных условиях. Результаты показали, что определенная группа клиентов — называемые «клиентами с непредвиденными обстоятельствами овердрафта», которые иногда перерасходуют средства, но не регулярно — в конечном итоге имеют шаблоны невыполнения обязательств, наиболее близкие к фактическим наблюдаемым данным, и, таким образом, представляют значительный риск. Авторы приходят к выводу, что банки

должны уделять особое внимание этим клиентам со средним уровнем риска, чтобы лучше управлять рисками портфеля кредитных карт.

MAS также использовались для разработки систем поддержки принятия решений по кредитному риску. Например, исследователи предложили многоагентные архитектуры для кредитного скоринга и оценки кредита, где разные агенты могут оценивать различные аспекты заявки на кредит (один агент проверяет кредитную историю, другой оценивает рыночные условия и т. д.), а затем коллективно принимать решение об одобрении или отклонении кредита. Такие системы могут быть адаптивными, обучаясь на результатах, чтобы со временем улучшать кредитные оценки. Хотя эти архитектуры все еще в значительной степени экспериментальны, они указывают на потенциал MAS для дополнения традиционных инструментов кредитного риска путем добавления слоя распределенного интеллекта и перекрестной проверки среди агентов.

Помимо индивидуальных портфелей, MAS широко применялись к системному кредитному риску, где существует опасение, что беда в одном учреждении (или группе заемщиков) может распространиться по финансовой сети. В этих моделях банки часто являются агентами, а связи между агентами представляют кредитные риски (такие как межбанковские кредиты, позиции по дериватам или общие активы). Цель состоит в том, чтобы смоделировать, как шок — например, дефолт крупного контрагента или внезапная потеря стоимости активов — может привести к каскадным сбоям или системному кризису.

Модель ABBA (Agent-Based Banking Analysis) МВФ, которая включает вкладчиков, заемщиков и банки в качестве взаимодействующих агентов. В ABBA банки кредитуют фирмы и друг друга, и они должны удовлетворять нормативным требованиям к капиталу и ликвидности. Моделирование с использованием ABBA использовалось для изучения таких сценариев, как изменения в правилах банковского капитала или шоки финансирования. Например, если требования к капиталу установлены слишком низкими по отношению к базовому кредитному риску, модель может показать накопление кредитования и взаимосвязанных рисков, что потенциально приводит к большему количеству банковских банкротств, когда наступает кризис. С другой стороны, более высокие требования к капиталу или ликвидности могут сократить количество неудач, но также могут ограничить предложение кредита — компромисс, который ABBA позволяет аналитикам подробно исследовать. Эти выводы на основе агентов дополняют традиционные стресс-тесты, включая адаптацию поведения: банки в ABBA будут

пытаться перебалансировать портфели или сохранить капитал, сокращая дивиденды в условиях стресса, который отражает действия в реальном мире.

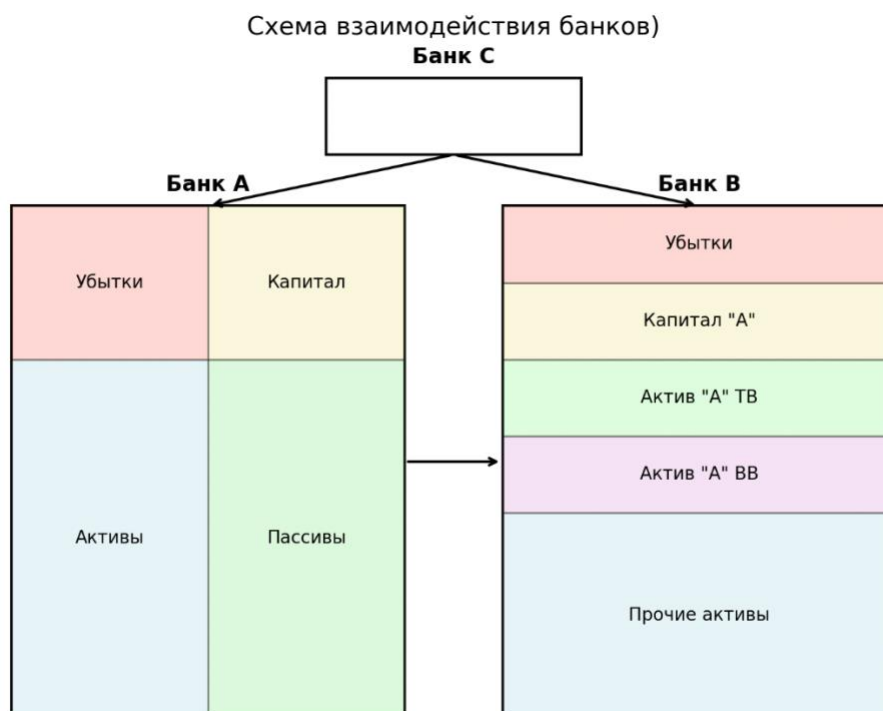


Рисунок 16. Распространения убытков в межбанковской сети на основе схемы. Шок активов банка А (красные убытки на балансе) уменьшает его капитал и приводит к списанию активов, удерживаемых банком В (требования к банку А) и банком С. Стрелки показывают, как убытки одного банка передаются другим; например, убытки банка В, в свою очередь, влияют на банк С (который имеет требования к В), а убытки банка С возвращаются в банк А. Такие агентные симуляции помогают визуализировать пути распространения кредитного риска в банковской системе.

Модели MAS подходят для экспериментов с политикой. Например, можно добавить нового агента, представляющего регулятора, который устанавливает определенные правила (например, ограничение на соотношение кредитов к депозитам или механизм спасения), а затем провести симуляцию, чтобы увидеть, насколько это снижает риск. Исследования действительно использовали АВМ для тестирования макропруденциальной политики, такой как контрциклические буферы капитала или налоги на межбанковское кредитование. Возможность переключать правила и видеть результаты делает MAS привлекательным инструментом для изучения влияния нормативных актов в виртуальной среде перед их реализацией в реальности.

Однако существуют также проблемы и ограничения. Модели MAS могут быть жадными до данных — если быть реалистичными, они часто требуют подробных микроданных (например, распределение доходов заемщиков, сеть межбанковских рисков и т. д.). В некоторых случаях эти данные являются собственностью или их нелегко получить. Это может заставить разработчиков моделей делать упрощающие предположения, которые могут снизить точность. Еще одной проблемой является коммуникация и принятие этих моделей. Их сложность может сделать их «черными ящиками» для лиц, принимающих решения, особенно если они не знакомы с подходами к моделированию. Обеспечение прозрачности (посредством четкой документации и, возможно, более простых сводных версий модели) является ключевым фактором, если MAS будут использоваться в реальном управлении банковскими рисками или регулирующими органами.

С вычислительной точки зрения, крупномасштабные симуляции могут столкнуться с проблемами производительности, но с современными вычислениями это уже не такое узкое место, как раньше. Тем не менее, если кто-то захочет смоделировать каждый кредит и каждого заемщика в большой экономике, это могут быть миллионы агентов, что потребует либо массивных параллельных вычислений, либо умной агрегации в модели. Исследователи часто находят золотую середину, моделируя репрезентативные группы агентов, а не каждого отдельного человека (например, используя несколько тысяч «кластеров» агентов для представления миллионов похожих заемщиков).

Необходимо также учитывать, что MAS хороши настолько, насколько хороши их правила — если агенты не запрограммированы на реалистичное поведение, результаты не будут информативными. Таким образом, для создания надежных правил агентов (например, как именно ведут себя кредитные специалисты или как банки управляют ликвидностью) необходимы экспертные знания в области банковского дела. Междисциплинарное сотрудничество между экономистами, специалистами по информатике и риск-менеджерами полезно по этой причине.

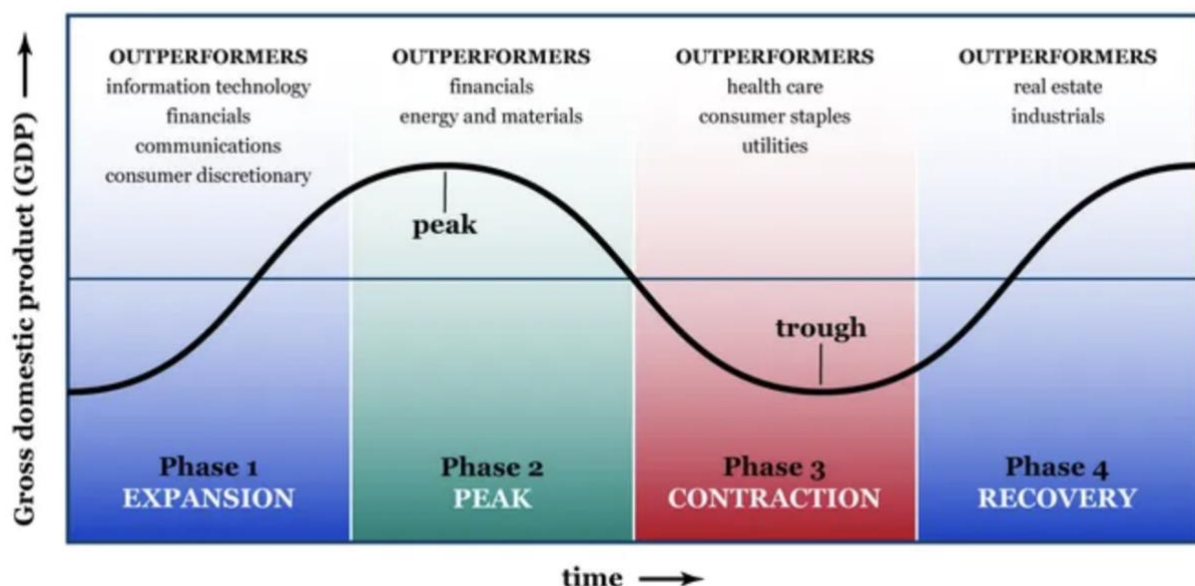
Подводя итог, можно сказать, что моделирование кредитного риска на основе MAS показало большие перспективы в улучшении нашего понимания как идиосинкразического, так и системного риска в банковской сфере. Оно согласуется с растущим признанием того, что финансовые системы являются сложными адаптивными системами, а не статическими машинами. Принимая эту сложность посредством моделирования, мы можем получить предвидение потенциальных уязвимостей и проверить эффективность мер по снижению риска в безопасной среде. Сейчас обсуждение

идет не столько о том, следует ли использовать MAS, сколько о том, как интегрировать их в инструментарий банков и регуляторов наряду с традиционными моделями, используя сильные стороны каждой из них.

Глава 3: Экономический цикл и кризисные ситуации

3.1 Фазы экономического цикла: рост, пик, спад, дно и восстановление

Four phases of an economic cycle



© Encyclopedia Britannica, Inc.

Рисунок 17. Все этапы экономического цикла – рост, пик, рецессия и дно.

Во время фазы роста объем производства начинает расти. Предприятия начинают инвестировать в новые проекты; найм увеличивается, занятость и доходы растут. Кредиты легко доступны в банках по относительно низким процентным ставкам, что еще сильнее стимулирует потребительские расходы и бизнес-инвестиции. Инфляция может быть низкой или умеренной, и по мере усиления спроса цены могут начать расти.

Формула для расчета роста ВВП:

$$g = \frac{\text{ВВП}_i - \text{ВВП}_{i-1}}{\text{ВВП}_{i-1}}$$

После роста следует фаза пика. Пик характеризуется тем, что экономика работает на уровне или выше своих возможностей. Продажи и прибыли предприятий достигают своего пика, а безработица падает до очень низкого уровня. Пик – прекрасное время для появления финансовых пузырей, то есть цены растут нереалистично из-за чрезмерных

ожиданий. Процентные ставки тоже могут расти, поскольку центральный банк пытается охладить перегретую экономику. Пик по сути является поворотным моментом – рост достигает своего максимального уровня и начинает замедляться. Предприятиям может быть трудно расширяться дальше, и некоторые заемщики становятся слишком аккредитованными.

После пика экономика вступает в рецессию. Рецессия определяется как значительное снижение экономической активности, обычно заметное по падению ВВП, снижению доходов, росту безработицы и снижению производства в течение нескольких кварталов. Во время рецессии потребительские расходы падают; предприятия сокращают выпуск и инвестиции из-за слабого спроса. Корпоративная прибыль сокращается, а некоторые фирмы терпят убытки, что приводит к увольнениям и росту безработицы. Кредитные условия ужесточаются – банки становятся более осторожными или сами сталкиваются с убытками, поэтому кредитование замедляется. Компании с большой задолженностью могут объявить дефолт; цены на активы (акции, жилье и т. д.) часто резко падают. Таким образом происходит сокращение неэффективных инвестиций, и корректировка цен в сторону понижения. Рецессии могут быть мягкими или жесткими (финансовый кризис). Политики обычно реагируют смягчением денежно-кредитной политики (снижением процентных ставок), а иногда и увеличением государственных расходов, чтобы смягчить спад.

В конце концов, рецессия достигает дна. В этот момент экономическая активность падает до самого низкого уровня. В фазе восстановления ВВП перестает сокращаться и снова начинает расти. Компании, которые выжили, начинают стабилизировать продажи; начинается «осторожное» инвестирование. Безработица, которая достигла пика во время рецессии, снова начинает снижаться, хотя это может быть запаздывающим показателем (рабочие места восстанавливаются медленнее, чем объемы производства). Уверенность возвращается небольшими шагами – банки начинают немного больше кредитовать, а потребители увеличивают расходы. Часто государственная поддержка во время рецессии (например, программы стимулирования или меры по спасению банков) помогает ускорить восстановление. По мере укрепления рынка начинается фаза роста, тем самым экономический цикл повторяется.

Важно отметить, что продолжительность и интенсивность каждой фазы может отличаться. Некоторые подъемы продолжительны и постоянны, в то время как другие могут быть быстрыми и кратковременными. Рецессия также может варьироваться от краткой и

неглубокой до продолжительной. На экономический цикл влияют многие факторы — технологические изменения, политические решения, внешние потрясения (например, скачки цен на нефть или пандемия) и, как подчеркнул Хайман Мински, поведение финансовых рынков.

Во время подъема и пика ключевые макроэкономические показатели движутся характерными способами. Например, рост ВВП силен во время подъемов и слаб или отрицателен во время рецессий; безработица падает во время подъемов и растет во время спадов; уровни кредитования и задолженности обычно растут во время подъемов и могут сокращаться во время спадов. На рисунке 18 показан пример темпов роста ВВП для цикла подъема-спада в двух странах (США и Россия) во время кризиса 2008 года, подчеркивая общую картину резкого падения после пикового года.

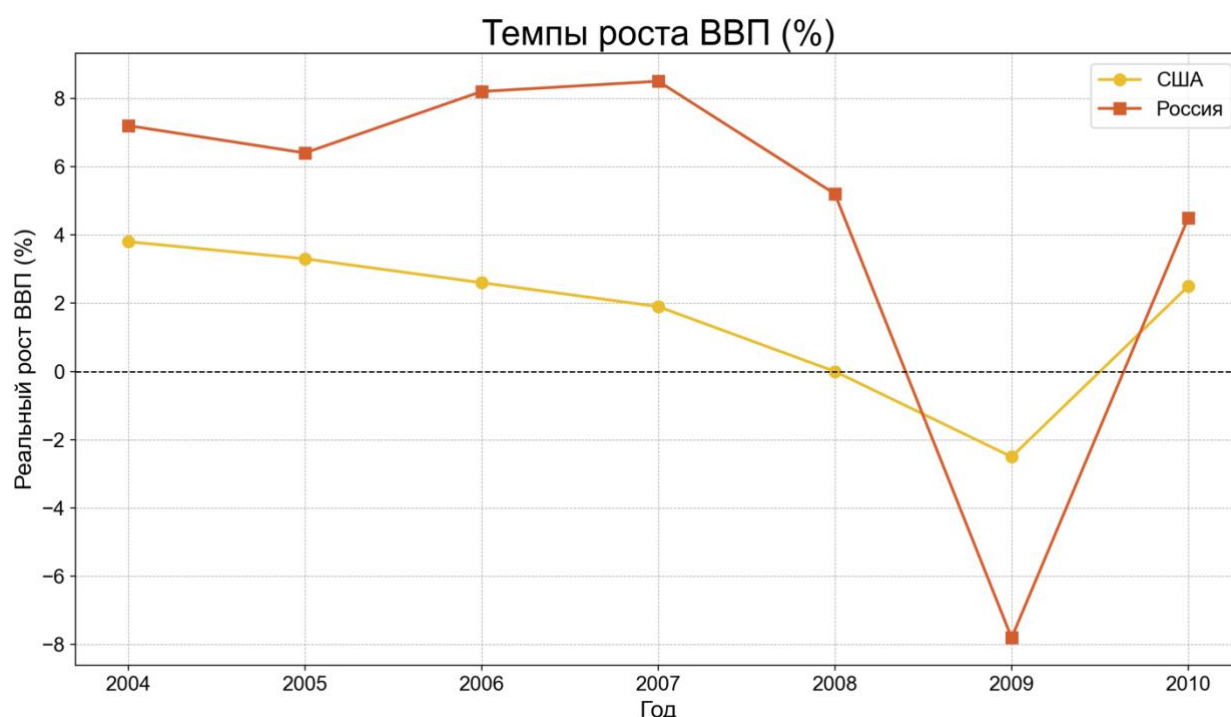


Рисунок 18. Темпы роста ВВП (%) для США и России, 2004–2010 гг. Обе страны демонстрировали сильный рост до 2008 г. (особенно российский нефтяной бум) далее мировой финансовый кризис привел к резкому спаду: в 2009 г. ВВП США упал примерно на 2,5%, а ВВП России — примерно на 7,8%. Восстановление началось в 2010 г.

Для более детального анализа экономического цикла я воспользуюсь я воспользуюсь теорией Хаймана Мински. Мински описал процесс, в котором длительный период

стабильных «экономических времен» (рост, низкий уровень дефолта, растущие цены на активы) побуждает экономических субъектов брать на себя все больше риска и долгов. Банки, фирмы и домохозяйства становятся более оптимистичными и увеличивают свои инвестиции, предполагая, что такой благоприятный период времени будет продолжаться еще очень долго. Этот прогрессивный рост задолженности и риска можно рассматривать в трех стадиях финансирования, которые Мински и обрисовал.

1. В начале посткризисного периода или в период стабильности заемщики и кредиторы проявляют осторожность. Представители хедж-финансирования — это те, кто берет займы в расчете на то, что их будущих доходов или денежных потоков будет достаточно для выплаты процентов и основной суммы долга по займам. Другими словами, долги принимаются консервативно, и заемщик может полностью «застраховаться» от рисков, покрывая все обязательства за счет нормального дохода. Например, фирма, которая берет кредит и планирует выплачивать его из своей прибыли в течение нескольких лет, или покупатель жилья, который берет ипотеку, которую можно постепенно выплачивать из своей зарплаты, — это примеры хедж-финансирования. На данном этапе балансовые отчеты, как правило, надежны. Общий уровень левериджа (соотношение долга к собственному капиталу) в экономике умеренный.
2. По мере роста оптимизма во время экономического роста как заемщики, так и кредиторы готовы идти на больший риск. Представители спекулятивного финансирования — это те, кто все еще может выплачивать проценты по своему долгу из своего дохода, но не может выплачивать основную сумму долга — они должны рефинансировать долг. Эти заемщики делают ставку на то, что они смогут рефинансировать свои займы до тех пор, пока кредиты остаются доступными. Примером может служить застройщик, который берет краткосрочный кредит на строительство объекта, проценты по кредиту могут быть выплачены из дохода от аренды, но основная сумма будет выплачена только за счет получения нового кредита или продажи недвижимости по более высокой цене. На этом спекулятивном этапе долги растут быстрее, чем доходы от основной деятельности. Коэффициент использования заемных средств в экономике увеличивается. Кредитные рынки внешне остаются стабильными, но зависимость от рефинансирования делает их уязвимыми — если кредиторы внезапно откажутся от своих обязательств или процентные ставки вырастут, у этих спекулятивных подразделений могут возникнуть проблемы с выполнением своих обязательств.

Тем не менее, во время спекулятивной фазы цены на активы часто продолжают расти (акции, недвижимость и т. д.), и это стимулирует еще большее заимствование.

3. В самый разгар экономического бума некоторые заемщики прибегают к «финансам Понци» (названной в честь печально известной схемы Чарльза Понци). Они не могут полностью выплачивать проценты или основную сумму долга из своего текущего дохода. Единственный способ сохранить свое положение — это занимать больше средств или продавать активы по постоянно более высоким ценам. Они полагаются на постоянно растущую стоимость активов, чтобы оправдать свои займы. Например, рассмотрим человека во время жилищного кризиса, который берет кредит на недвижимость (ипотека) с очень низкой начальной процентной, исходя из предположения, что цены на жилье будут продолжать расти. Такая стратегия работает только в том случае, если цены на активы продолжают расти, а кредит легко доступен.

Важнейшее открытие Мински заключалось в том, что, когда экономика проходит через эти стадии экономического роста, это неизбежно создает условия для кризиса.

Длительный период стабильности (небольшое количество дефолтов, устойчивый экономический рост) приводит к снижению воспринимаемого риска – кредиторы ослабляют кредитные стандарты, заемщики получают больше заемных средств.

Изначально безопасное хеджевое финансирование уступает место более рискованному спекулятивному финансированию и в итоге превращается в финансовую пирамиду для определенного сегмента экономики. На этом этапе система чрезвычайно хрупка:

малейший толчок может привести к резкому изменению ситуации, когда участники поймут, что император голый. Этот момент резкого разворота часто называют “моментом Мински”. Это момент, когда доверие рушится: кредиторы внезапно отказываются пролонгировать кредиты или требуют гораздо более высокого обеспечения; цены на активы падают, потому что все пытаются их продать; и многие заемщики, которые на бумаге казались прибыльными, оказываются неплатежеспособными. За этим следует каскадный финансовый кризис – по сути, пик, переходящий в крах. Банки несут большие убытки по безнадежным кредитам, инвесторы устремляются к более безопасным активам, кредиты иссякают даже для здоровых предприятий, а реальная экономика испытывает нехватку финансирования, что усугубляет экономический спад. Другими словами, гипотеза финансовой нестабильности объясняет, как экономический бум, подпитываемый долгом, может практически за одну ночь превратиться в кризис. Стоит отметить, что

теория Мински противоречит более традиционным взглядам, которые часто рассматривают циклы как результат внешних потрясений или политических ошибок. Вместо этого Мински делает акцент на внутренней динамике капиталистических экономик: сам успех стабильного периода порождает поведение, подрывающее стабильность. Долгое время эта точка зрения рассматривалась за рамками мейнстрима – до 2008 года многие экономисты и политики полагали, что передовые финансовые системы стали практически устойчивыми к кризисам благодаря сложному управлению рисками и регулированию. Кризис 2008 года доказал обратное, и внезапно предупреждения Мински, на которые раньше не обращали внимания, оказались пророческими.

3.2 Финансовый кризис 2008 года

Финансовый кризис 2008 года (часто называемый глобальным финансовым кризисом) стал одним из самых серьезных экономических спадов со времен Великой депрессии. Это служит наглядным примером описанного мною цикла: бурный экономический рост, достигший своего пика и внезапно перешедший в глубокую рецессию, за которой последовало восстановление только после масштабного политического вмешательства.

В середине 2000-х годов в Соединенных Штатах (и ряде других стран) наблюдался сильный экономический рост. В США это характеризовалось бурным развитием рынка жилья, низкими процентными ставками и удобными условиями кредитования. Несколько ключевых факторов способствовали росту рисков в этот период:

1. Цены на жилье в США резко выросли в период с 2000 по 2006 год, чему способствовали легкие ипотечные кредиты и спекулятивный ажиотаж. Банки и ипотечные кредиторы предоставили кредиты миллионам американцев на покупку жилья, часто требуя очень низких первоначальных платежей. Широкое распространение получили новые финансовые продукты, в частности ссуды заемщикам с плохой кредитной историей. Как и в случае с Минском, стандарты кредитования со временем ухудшились - в начале 2000-х годов для большинства кредитов по-прежнему требовались документы о доходах и разумное соотношение суммы кредита к стоимости (хедж-финансирование). Но к моменту расцвета «финансирования Понци» стало обычным делом: многие покупатели жилья брали кредиты, полагая, что смогут быстро рефинансировать или продать с прибылью, а

не возвращать их за счет своих доходов. К 2006 году стоимость жилья была намного выше уровня, оправданного доходами от аренды. По сути, рынок жилья стал полагаться на постоянно растущие цены, чтобы поддерживать себя.

2. Банковская и теневая банковская система получила высокую долю заемных средств и активно инвестировала в жилищный бум. Крупные финансовые учреждения создавали ценные бумаги, обеспеченные ипотекой (MBS), и долговые обязательства, обеспеченные залогом (CDO), которые объединяли ипотечные кредиты и якобы распределяли риски. Какое-то время казалось, что это нововведение сделало систему более безопасной - число дефолтов было низким, и инвесторы по всему миру стремились покупать эти высокодоходные ценные бумаги, полагая, что они почти так же безопасны, как государственные облигации. Однако эти продукты фактически концентрировали риски рынка жилья в финансовой системе. К 2007 году у крупных инвестиционных банков, таких как Lehman Brothers, Bear Stearns и Merrill Lynch, соотношение заемных средств составляло порядка 30:1, что означало, что они были чрезвычайно уязвимы даже к небольшому падению стоимости активов
3. Одной из причин столь высокого уровня заемных средств и низкого качества кредитования является вера в то, что «рынки знают лучше». Финансовое дерегулирование в предыдущие десятилетия (например, отмена закона Гласса-Стигалла в 1999 году, который разделял коммерческую и инвестиционную банковскую деятельность) и легкий подход к деривативам позволили банкам заниматься более рискованной деятельностью без надлежащего надзора. Кредитные рейтинговые агентства присвоили чрезмерно оптимистичные рейтинги ипотечным продуктам, недооценив риск массового дефолта. Этот широко распространенный оптимизм является именно тем, что предсказывает теория Мински перед кризисом: экономические агенты коллективно недооценивают риск, потому что в недавнем прошлом не было серьезных кризисов. Фактически, вплоть до 2008 года было распространено предположение, что цены на жилье в США по всей стране существенно не упадут, поскольку этого не происходило уже много десятилетий. Это предположение оказалось неверным.

Настоящий «момент Мински» наступил в 2007–2008 годах, когда стало ясно, что ипотечные активы гораздо более рискованны, чем считалось ранее. Несколько событий стали причиной краха:

1. В 2007 году ряд ипотечных кредиторов в США обанкротились из-за роста числа неплатежей по кредитам.
2. Банки начали опасаться за платежеспособность друг друга, что привело к резкому росту ставок по межбанковским кредитам. Известный показатель, спред TED (разница между процентной ставкой по межбанковским кредитам и краткосрочным государственным долгом США), резко подскочил в 2007–2008 годах, что отражает стресс в банковской системе. Ликвидность начала иссякать, что является отличительной чертой минского кризиса, когда все отказываются от кредитования.
3. Кризис усугубился в сентябре 2008 года, когда обанкротились крупнейшие банки. Известный инвестиционный банк Lehman Brothers объявил о банкротстве после того, как он больше не мог выплачивать свои краткосрочные долги. Примерно в то же время AIG, огромная страховая компания, которая заключила кредитные дефолтные свопы на триллионы долларов (по сути, страхование рискованных ипотечных облигаций), балансировала на грани краха, и правительству пришлось ее спасать. Фондовый рынок резко упал (индекс S&P 500 потерял около 50% своего значения с пика 2007 года к началу 2009 года), а кредитные рынки замерли из-за начавшейся паники.

3.3 Последствия для банковского сектора США

Непосредственные последствия кризиса в Соединенных Штатах были наиболее заметны в банковском и финансовом секторе. Поскольку цены на жилье упали, а число неплатежей по ипотечным кредитам резко возросло, финансовые учреждения понесли огромные убытки по активам, обеспеченным ипотекой. Основные последствия для банков США включали:

1. Десятки банков обанкротились, и еще больше было спасено.
2. Крупным и малым предприятиям стало гораздо труднее получать кредиты; потребители столкнулись с ужесточением условий по ипотечным кредитам, автокредитам и кредитным картам. Фактически, объем банковского кредитования в США сократился самыми быстрыми темпами за период после Второй мировой войны после 2008 года, когда банки сократили долю заемных средств.
3. Кризис также высветил риски, связанные с «теневой банковской системой» – финансовыми фирмами, которые действуют как банки (берут краткосрочные займы для долгосрочного инвестирования), но не регулируются как банки. В эту

категорию попали инвестиционные банки, ипотечные кредиторы, фонды денежного рынка и страховые компании.

4. Чтобы поддержать банки и экономику, Федеральная резервная система не только снизила свою базовую процентную ставку фактически до 0% к концу 2008 года, но и приступила к количественному смягчению - покупке долгосрочных облигаций и ипотечных ценных бумаг для вливания ликвидности.

Несмотря на эти серьезные последствия, к середине 2009 года «кровотечение» в финансовом секторе США начало замедляться благодаря агрессивным политическим мерам реагирования. Критический момент наступил после проведения в 2009 году регулируемыми органами США банковских «стресс-тестов», в ходе которых проверялось, обладают ли крупные банки достаточным капиталом, чтобы выдержать наихудшие экономические сценарии. Обнародование результатов стресс-тестов в мае 2009 года, а также привлечение банками дополнительного капитала помогли восстановить доверие. Постепенно крупные банки США стабилизировались. Этот эпизод привел к крупным реформам в США, в частности к принятию Закона Додда-Франка (2010), который ужесточил финансовое регулирование, включая более высокие требования к капиталу банков, надзор за системными небанковскими учреждениями и меры по защите прав потребителей. Федеральная резервная система также получила расширенные полномочия по мониторингу системных рисков. Эти изменения были направлены на устранение уязвимостей, на которые указывает теория Мински, а именно на предотвращение неконтролируемого наращивания заемных средств и рискованного кредитования во время бумов, которые могут привести к серьезным спадам.

3.4 Последствия для российского банковского сектора

Кризис 2008 года, начавшийся в Соединенных Штатах, быстро распространился по всему миру. Россия, как и многие страны с формирующейся рыночной экономикой, переживала сильный экономический подъем, во многом обусловленный высокими ценами на сырьевые товары (особенно на нефть) и значительным притоком капитала. Когда мировая финансовая ситуация изменилась, Россия тоже пострадала, хотя немного иначе, чем США. В 2008–2009 годах последствия для российских банков и экономики были значительными:

1. Российские эксперты часто описывают последствия кризиса как «двойной шок» для России: во-первых, финансовый шок, вызванный ужесточением условий кредитования в мире, и, во-вторых, шок цен на сырьевые товары, вызванный падением цен на нефть (ключевую статью российского экспорта). В середине 2008 года цены на нефть превышали 140 долларов за баррель; к началу 2009 года они упали до уровня ниже 40 долларов. В отличие от США, где кризис начался в сфере жилищного строительства, проблемы России были в значительной степени импортированы – результат интеграции в мировые финансовые рынки и торговлю сырьевыми товарами.
2. ВВП России, который в середине 2000-х годов рос на 7–8% в год, в 2009 году сократился примерно на 7,8%, что стало одним из худших показателей снижения среди экономик G20. Промышленное производство упало более чем на 10%, а безработица и бедность заметно выросли, поскольку шахты, заводы и строительные проекты остановились или замедлились. Для банковского сектора наиболее явным признаком стресса стал всплеск проблемных кредитов. Во время бума российские банки быстро наращивали объемы кредитования (зачастую объем выданных кредитов увеличивался более чем на 30% в год). Когда разразился кризис, многие заемщики – как предприятия, так и потребители – не смогли обслуживать свои долги.
3. В последнем квартале 2008 года, когда мировые кредитные рынки замерли, российские банки столкнулись с серьезным дефицитом ликвидности. Банки, которые полагались на краткосрочное оптовое финансирование (часто с зарубежных рынков), внезапно оказались отрезанными. Ситуация с межбанковским кредитованием в России также осложнилась, поскольку банки стали опасаться за стабильность друг друга. Российские власти предприняли ряд мер для обеспечения ликвидности и стабилизации банковской системы. Центральный банк России (ЦБ РФ) быстро снизил процентные ставки и предоставил банкам рублевую ликвидность. К счастью, полномасштабного банкротства банков удалось избежать, и депозиты начали возвращаться к началу 2009 года после того, как власти продемонстрировали готовность поддержать банки.
4. Кризис также ударил по российской валюте. В конце 2008 года рубль оказался под сильным давлением, поскольку инвесторы выводили средства. Центральный банк израсходовал значительную часть своих резервов (более 200 миллиардов долларов, что составляет около трети резервных запасов), пытаясь защитить курс рубля с

помощью интервенций. В конечном счете они позволили провести управляемую девальвацию: обменный курс рубля был постепенно ослаблен, чтобы отразить новую реальность снижения цен на нефть и оттока капитала. Ослабление рубля оказало неоднозначное воздействие – оно частично смягчило удар по отечественной промышленности (сделав российские товары немного более конкурентоспособными), но также повысило стоимость обслуживания долгов в иностранной валюте для российских банков и компаний. Во время экономического бума многие российские банки и фирмы брали кредиты в долларах или евро (поскольку ставки за рубежом были ниже).

Кризис в США был вызван внутренними финансовыми потрясениями (лопнувший кредитный пузырь), в то время как кризис в России был в значительной степени вызван внешним воздействием (уход глобальных инвесторов и обвал цен на сырьевые товары), наложившимся на некоторые внутренние кредитные проблемы. США были вынуждены напрямую спасать банки, которые сделали неудачные инвестиции; Россия была вынуждена поддерживать банки в первую очередь для того, чтобы пережить внезапное прекращение внешнего финансирования и экономический спад. В США эпицентром кризиса стало финансирование жилищного строительства; в России это было больше связано с корпоративным долгом и макроэкономическим шоком. Тем не менее, в обеих странах мы наблюдаем общие закономерности, согласующиеся с теорией циклов: период бума, сопровождающийся ростом заемных средств и доверия, за которым следует внезапный перелом и массированное вмешательство правительства для стабилизации системы и начала восстановления.

3.5 Выводы после кризиса

К середине 2009 года и США, и Россия, да и вообще большая часть мира, предпринимали шаги по преодолению экономического спада и переходу к фазе восстановления. Экономическому подъему в США способствовала стимулирующая политика, в России экономическому подъему способствовало повышение цен на сырьевые товары и восстановление доверия благодаря оперативному антикризисному управлению. Однако последствия кризиса оставались, а последовавший за ним экономический рост был более сдержанным, чем предшествовавший ему бум. Выводы, которые были сделаны мировыми сообществами:

1. Кризис показал, что одного макроэкономического роста недостаточно – важны качество и устойчивость. Кредитный бум, который опережает фундаментальные показатели экономики, скорее всего, закончится кризисом. Политики во всем мире признали необходимость улучшения надзора за финансовой системой. Это привело к созданию новых или усилению старых регулирующих органов (например, в США создан совет по надзору за финансовой стабильностью для мониторинга системных рисков) и международные реформы (соглашения Базель III, которые ввели более высокие требования к капиталу и ликвидности для банков по всему миру). Циклы экономического роста и рецессии не могут быть устранены, но более эффективная макропруденциальная политика потенциально может смягчить крайности финансового цикла.
2. Эпизод 2008 года подтвердил давно игнорируемое мнение Хаймана Мински о том, что стабильность порождает нестабильность. Во время посткризисного самоанализа работы Мински стали обязательными к прочтению. Термин “момент Мински” вошел в популярный лексикон для описания переломных моментов кризисов. Стало понятно, что власти должны следить за признаками спекулятивного избытка, такими как растущее соотношение долга к доходам, расхождение цен на активы с фундаментальными показателями или быстрое распространение новых сложных финансовых продуктов, и принимать превентивные меры.
3. Контраст между опытом США и России также подчеркивает, насколько взаимосвязана мировая экономика. Финансовый шок в одной стране может передаваться через потоки капитала, торговлю и каналы доверия. Таким образом, важно поддерживать надежные международные системы социальной защиты. Пример России показал, что даже страна с, казалось бы, прочными фундаментальными показателями (низкий государственный долг, профицит бюджета, большие резервы) может серьезно пострадать от глобального кризиса, если ее экономика не диверсифицирована, а ее банки и фирмы в значительной степени зависят от иностранного финансирования.

Глава 4: Прогнозирование банковского дефолта. Логистическая регрессия VS Случайный лес

4.1 Описание переменных

В своей работе я анализирую банковский дефолт ($\text{target} = 1$ в случае дефолта, 0 в противном случае), используя две модели – логистическую регрессию и классификатор случайных лесов. Далее подробно описаны переменные, которые будут в моделях:

1. Рост ВВП (%) – квартальный показатель роста валового внутреннего продукта, отражающий макроэкономические условия. Более высокие темпы роста ВВП свидетельствуют о сильной экономике, что в целом снижает риск дефолта, и наоборот, низкие или отрицательные темпы роста увеличивают риск дефолта.
2. Инфляция (%) – годовой уровень инфляции (в годовом исчислении). Я ожидаю, что более высокая инфляция будет коррелировать с более высоким риском дефолта, как это наблюдается в кризисные периоды.
3. Ключевая ставка (%) – ключевая процентная ставка центрального банка на конец квартала. Она влияет на стоимость финансирования и спрос на кредиты. Я использую квадратичный показатель (key_rate_sq) для учета нелинейных эффектов. Например, умеренные ставки могут быть благоприятными, но очень высокие ставки (жесткая денежно-кредитная политика) могут резко увеличить риск дефолта из-за повышения стоимости заимствований. Я ожидаю положительного коэффициента для key_rate и отрицательного коэффициента для key_rate_sq , что соответствует эффекту обратной U-образной формы (риск дефолта возрастает с повышением ставки до определенного уровня, затем стабилизируется или падает).
4. Достаточность капитала (N1) – коэффициент достаточности регулятивного капитала (капитал 1-го уровня в процентах от активов, взвешенных с учетом риска). Для линейности я использую логарифмическое значение N1 (\log_N1). Достаточный запас капитала защищает от убытков, поэтому низкий N1 является надежным предиктором банкротства. Мне кажется, что коэффициент \log_N1 будет отрицательным (более высокий капитал \rightarrow более низкая вероятность дефолта).
5. Процент проблемных кредитов (NPL %) – отношение проблемных кредитов к общему объему выданных кредитов. Это показатель качества активов; растущий

коэффициент проблемных кредитов сигнализирует об ухудшении кредитного портфеля. Увеличение количества проблемных кредитов является знаком бедственного положения банка. Таким образом, ожидается положительный коэффициент (больше проблемных кредитов → более высокий риск дефолта).

6. Рентабельность активов (ROA %) – отношение чистой прибыли к общей сумме активов, указывающее на прибыльность. Низкая или отрицательная рентабельность означает, что банк убыточен, часто из-за высоких резервов на возможные потери по ссудам или неэффективности. Снижение дохода предшествует банкротству, поэтому коэффициент рентабельности инвестиций должен быть отрицательным (более низкая прибыль → более высокий риск дефолта).
7. Коэффициент ликвидности (НЗ %) – нормативный показатель соотношения ликвидных активов и краткосрочных обязательств. НЗ указывает на способность банка противостоять оттоку депозитов или нехватке финансирования. Более высокий коэффициент ликвидности, как правило, является защитным (должен снижать риск дефолта). Мне кажется, что коэффициент для НЗ будет отрицательным (хотя в моей выборке эта переменная показала положительный знак, о чем будет рассказано ниже).
8. Размер банка (log_assets) – логарифмический показатель совокупных активов. Крупные банки могут быть более диверсифицированными и “слишком крупными, чтобы обанкротиться”, что подразумевает меньший риск, но они также могут принимать на себя больше рисков. Мы ожидаем, что коэффициент log_assets будет отрицательным (небольшие банки более склонны к дефолту, как это часто наблюдалось в прошлом).
9. Системная значимость (share_sys) – доля активов Банка в банковской системе (%). Более высокая доля указывает на системно важный банк. Такие банки могут получать государственную поддержку (снижающую риск дефолта), но если они и терпят крах, то часто во время системных кризисов.

Для модели случайного леса я включил все вышеперечисленные переменные и добавили еще пять функций, отражающих рост и структуру:

10. Рост кредитного портфеля (%) – темпы роста кредитного портфеля банка. Быстрый рост кредитования может предвещать проблемы с качеством активов в будущем. Ожидается, что высокий рост кредитования увеличит риск дефолта.

11. Соотношение кредитов к депозитам (LDR) - кредиты, разделенные на депозиты, указывает на зависимость от оптового финансирования. Высокий LDR означает, что банк пользуется ссудами и, возможно, использует неосновное финансирование, что повышает риск в случае сокращения ликвидности. Таким образом, более высокий LDR, вероятно, повышает риск дефолта.
12. Доля межбанковского финансирования (%) – доля финансирования банка за счет межбанковских кредитов. Сильная зависимость от межбанковского финансирования может быть рискованной; потеря доверия рынка может быстро привести к прекращению финансирования. Ожидается, что более высокая доля межбанковского финансирования увеличит риск дефолта.
13. Рост прибыли (%) – темпы роста прибыли банка (в годовом исчислении). Это дополняет показатель рентабельности инвестиций, отражая тенденцию: сокращение прибыли (отрицательный рост) может свидетельствовать о том, что банк склоняется к убыткам и дефолту.
14. Тип банка – категориальный показатель принадлежности/типа банка (например, государственный, частный, иностранный). Эти фиктивные переменные отражают структурные различия: например, государственные банки могут иметь неявные гарантии (более низкий риск), в то время как другие могут быть более рискованными. Использование фиктивных параметров `bank_type` позволяет Random Forest учитывать фиксированные эффекты определенных категорий банков.

Вместе эти переменные охватывают влияние макроэкономического цикла и финансовое состояние конкретного банка для прогнозирования дефолта.

4.2 Выбор модели и гипотеза

В своей работе я буду использовать две взаимодополняющие модели: логистическую регрессию и случайный лес. Каждая из них обладает определенными преимуществами:

1. Логистическая регрессия (Logit) – это модель, которая является стандартом прогнозирования дефолта благодаря своей прозрачности и простоте интерпретации. Регулирующие органы часто отдают предпочтение логистическим моделям, поскольку можно легко проверить статистическую значимость и понять влияние каждого фактора на вероятность дефолта. Однако простой логит

предполагает линейную границу принятия решений и может иметь проблемы со сложными взаимодействиями или нелинейными эффектами в данных.

2. Случайный лес (RF) – это совокупность деревьев решений, которые дают мощный нелинейный классификатор. Случайные леса обеспечивают более высокую точность прогнозирования, чем логистические модели. Они могут отражать пороговые эффекты (например, коэффициент достаточности капитала ниже критического уровня) и взаимодействия между переменными без явного указания. Недостатком является ограниченная интерпретируемость – модель, по сути, является “черным ящиком” по сравнению с четкими оценками коэффициентов в logit. Интерпретация должна основываться на показателях важности признаков или графиках частичной зависимости, а не на простых коэффициентах.

Логистическая регрессия предоставляет интерпретируемую модель: можно определить, какие показатели являются статистически значимыми и в каком направлении, что позволяет получить информацию, соответствующую финансовой теории (например, высокий уровень проблемных кредитов и низкий уровень капитала — это плохо).

Случайный лес служит эталоном эффективности прогнозирования: он может выявлять сложные закономерности и устойчив к выбросам и мультиколлинеарности. Предыдущие исследования показали, что древовидные модели, такие как RF, часто превосходят logit по точности раннего предупреждения.

Моя гипотеза заключается в том, что модель случайного леса превзойдет логистическую регрессию по точности прогнозирования, особенно в нестабильные экономические периоды. То есть, RF должен будет лучше отражать различные модели дефолтов на этапах бума и спада.

4.3 Процесс и результаты моделирования

Исходный набор банковских данных был предварительно обработан и очищен (код приведен в приложении). Были исправлены пропущенные или аномальные значения, а финансовые переменные с неравномерным распределением (например, общая сумма активов, коэффициент достаточности капитала) были преобразованы в логарифмические данные для стабилизации дисперсии. Первоначальный тестовый запуск моделей показал, что некоторым предикторам не хватает статистической значимости; такие переменные

были удалены для упрощения моделей. Все остальные характеристики были измерены в натуральных единицах (или логарифмических эквивалентах), чтобы расчетные коэффициенты можно было интерпретировать напрямую. На протяжении всего процесса было уделено внимание тому, чтобы в обучающей выборке для каждой модели был представлен класс меньшинства (банки-неплательщики).

Окончательная спецификация модели была получена в результате итеративного процесса выбора и преобразования переменных. Ключевые шаги включали в себя:

1. Удаление несущественных предикторов - первоначальные регрессии выявили несколько переменных с высокими значениями p ; они были исключены.
2. Преобразование переменных для учета нелинейности - совокупные активы и капитал 1-го уровня, были преобразованы в логарифмическом порядке, чтобы отразить уменьшение влияния на риск дефолта. Процентная ставка была включена в квадратичном выражении для моделирования потенциального квадратичного эффекта.
3. Были изучены корреляции и дисперсионные коэффициенты инфляции (VIF). Были тщательно изучены переменные с высокой степенью коллинеарности (например, инфляция и Рост ВВП), и были оставлены только те, которые имели более строгое теоретическое обоснование и статистическую поддержку. Квадратичный показатель для учетной ставки был сохранен, несмотря на корреляцию с уровнем, поскольку это существенно улучшило соответствие.
4. На каждом этапе я оценивал степень соответствия (логарифмическая вероятность, псевдо R^2 и отбрасывал или добавлял переменные в свете экономических соображений.

Результаты для исходного (необработанного) набора данных (уровни значимости обозначены * для 10%, ** для 5%, *** для 1%):

Parameter	coef	std err	z	P> z	Значимость
const	0.4979	0.826	0.603	0.547	-
Инфляция	-0.1453	0.053	-2.745	0.006	***
Темп прироста ВВП	0.0048	0.014	0.348	0.728	-
Ключевая ставка в квадрате	0.0090	0.003	3.104	0.002	***

Норматив достаточности капитала (H1) (log)	0.0663	0.180	0.369	0.712	-
Доля просроченных кредитов (NPL%)	0.2010	0.023	8.919	0.000	***
Рентабельность активов (ROA%)	-0.3538	0.084	-4.200	0.000	***
Коэффициент ликвидности (H3%)	0.0007	0.000	1.558	0.119	-
Размер банка (лог активов)	-0.1106	0.026	-4.292	0.000	***
Доля банка в активах системы	-0.0456	0.027	-1.691	0.091	*

Таблица 2. Коэффициенты логистической регрессии (обучающая выборка) со стандартными ошибками и p-значениями без обработки данных.

Результаты для обработанного набора данных (уровни значимости обозначены * для 10%, ** для 5%, *** для 1%):

Parameter	coef	std err	z	P> z	Значимость
const	-1.6305	0.003	11.394	0.000	***
Инфляция	0.1318	0.050	-2.644	0.008	***
Ключевая ставка в квадрате	0.0080	0.002	3.217	0.001	***
Доля просроченных кредитов (NPL%)	0.2019	0.023	8.893	0.000	***
Рентабельность активов (ROA%)	-0.3360	0.081	-4.157	0.000	***
Размер банка (лог активов)	-0.1106	0.023	-4.873	0.000	***
Доля банка в активах системы	-0.0460	0.028	-1.673	0.094	*

Таблица 3. Коэффициенты логистической регрессии (обучающая выборка) со стандартными ошибками и p-значениями с обработкой данных.

После очистки (удаления экстремальных выбросов и сильно коллинеарных предикторов) окончательная модель содержала аналогичный набор предикторов, но коэффициенты немного изменились:

Факторы, повышающие риск дефолта

1. Инфляция (коэффициент =0,1318, $p=0,008$): Положительный коэффициент означает, что более высокая инфляция повышает риск дефолта. Проще говоря, когда инфляция растет, банки чаще терпят крах.
2. Ключевая процентная ставка в квадрате (коэффициент полезного действия=0,0080, $p=0,001$, *): также положительная и значимая величина. Поскольку это квадрат процентной ставки высокие процентные ставки резко увеличивают риск дефолта.
3. Коэффициент проблемных кредитов (% , коэффициент неработающих кредитов=0,2019, $p < 0,001$, *): сильно положительный. Банк с более высокой долей проблемных кредитов имеет гораздо более высокий риск дефолта

Факторы, снижающие риск дефолта

1. Рентабельность активов (ROA%, коэффициент=-0,3360, $p < 0,001$, *): Отрицательный коэффициент означает, что более прибыльные банки (более высокая рентабельность активов) имеют меньший риск дефолта.
2. Размер банка (логарифм активов, коэффициент=-0,1106, $p < 0,001$, *) - крупные банки (с более высоким логарифмом активов) менее склонны к дефолту.
3. Доля банка в активах системы (коэффициент=-0,0460, $p=0,094$): Коэффициент отрицательный (предполагает, что увеличение доли рынка может снизить риск). Есть небольшая тенденция к тому, что банки с более крупной долей рынка немного реже совершают дефолты.

$$\log \frac{P(Default = 1)}{P(Default = 0)} = -1.6305 - 0.1318 \cdot Inflation + 0.008 \cdot KeyRate^2 + 0.2019 \cdot NPL \\ - 0.3360 \cdot ROA + 0.0238 \cdot H3 - 0.1106 \cdot \ln(Assets) - 0.0460 \cdot share_sys$$

Я обучил классификатор случайных лесов на основе тех же данных (включая пять дополнительных переменных. В RF использовалось 100 деревьев принятия решений ($n_estimators=100$) с параметрами по умолчанию (выборка начальной загрузки, критерий

примеси Джини). Не было установлено ограничение на обрезку или глубину, что позволило лесу полностью соответствовать взаимодействиям. Я использовал стратифицированную выборку (то есть класс меньшинства был представлен в выборке каждого дерева).

1. Коэффициент неработающих кредитов (21,8%) — это, безусловно, главный показатель, который можно прогнозировать. Высокий коэффициент неработающих кредитов означает, что большая часть банковских кредитов находится в состоянии дефолта или проблемных. С экономической точки зрения это указывает на низкое качество активов и создание резервов на возможные потери.
2. Рентабельность активов (ROA%, 15,3%) – показатель рентабельности инвестиций, который измеряет прибыльность. Более высокая рентабельность инвестиций означает, что банк получает солидную прибыль от своих активов, что способствует наращиванию внутреннего капитала. Более прибыльные банки могут удерживать прибыль для покрытия будущих убытков и укрепления своих.
3. Рост прибыли банка в годовом исчислении (11,5%) – это свидетельствует о динамике роста прибыли. Если рост прибыли является устойчивым (увеличение прибыли в годовом исчислении), это свидетельствует об улучшении бизнес-показателей и увеличении резерва капитала.
4. Коэффициент достаточности капитала (Н1, логарифмический) (7,6%) – измеряет величину собственного капитала банка или капитала, способного покрыть убытки, по отношению к его активам, взвешенным с учетом риска. Более высокий уровень достаточности капитала означает, что банк имеет большую систему защиты.
5. Рост кредитного портфеля в годовом исчислении (7,6%) – показатель того, насколько быстро банк расширяет свой кредитный портфель. Умеренный рост кредитования может свидетельствовать о здоровом спросе и растущем бизнесе.
6. Доля межбанковских заимствований (7,3%) – это показатель того, какая часть финансирования банка поступает от других банков (оптовые краткосрочные заимствования). Высокая доля означает, что банк в значительной степени полагается на межбанковские рынки, а не на стабильные депозиты или собственный капитал.
7. Размер банка (логарифмические активы, 6,6%) – более крупные банки (по размеру активов) часто выигрывают от статуса “слишком крупного, чтобы обанкротиться” и диверсифицированных операций, что снижает вероятность дефолта.

8. Соотношение кредитов к депозитам (LDR%, 6,3%) - более высокий LDR означает, что банк выдал большую часть своих депозитов, оставив меньший запас ликвидности. Чем выше LDR, тем более рискованной (и потенциально более прибыльной) является кредитная стратегия банка. Чем ниже LDR, тем более безопасным (и менее прибыльным) является его подход.
9. Коэффициент ликвидности (НЗ%, 6,1%) – это показатель соотношения ликвидных активов (наличных денег, краткосрочных ценных бумаг и т. д.) к обязательствам. Высокий коэффициент ликвидности означает, что у банка достаточно наличных средств для покрытия внезапных изъятий средств; низкий коэффициент означает, что ему может быть сложно быстро привлечь средства.
10. Доля банка в системных активах (4,7%) – это рыночная доля банка в общем объеме активов отрасли. Более высокая доля означает, что банк является системно крупным.
11. Инфляция (3,1%) – высокая инфляция может снизить реальные доходы и увеличить неопределенность, что, как правило, повышает риск дефолта.
12. Ключевая ставка в квадрате (1,3%) – отражает влияние ключевой процентной ставки центрального банка (в квадрате, чтобы обеспечить нелинейный эффект). Повышение процентных ставок в итоге приводит к удорожанию заимствований и может замедлить экономический рост; очень высокие процентные ставки могут вызвать стрессовую ситуацию с кредитами (поскольку заемщики не выполняют свои обязательства по дорогостоящим долгам), а также снизить рыночную стоимость активов с фиксированной процентной ставкой.
13. Темпы роста ВВП (0,8%) – в данном случае рост ВВП является самым слабым показателем, отражающим общие экономические условия. Сильный рост ВВП обычно свидетельствует о здоровой экономике, в которой заемщики могут выплачивать долги, сокращая количество банковских дефолтов; и наоборот.

Пример дерева из Random Forest (глубина 3)

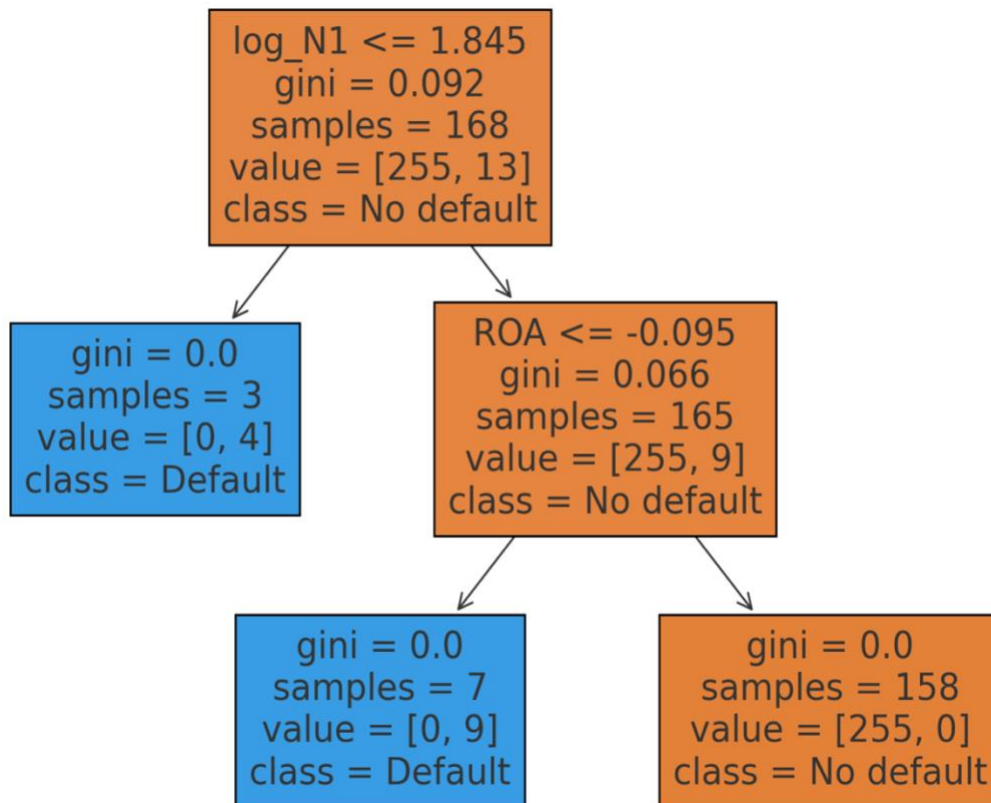


Рисунок 19. Оранжевые узлы указывают на класс большинства “Без дефолта”, синие узлы указывают на класс “По умолчанию”. Первое разбиение производится по достаточности капитала последующие разбиения связаны с ROA. В каждом узле показаны примеси Джини и количество образцов.

В этом дереве материнский узел разделяется на $\log_N1 \leq 1,845$. Это советует $\approx 6,33\%$. Если коэффициент достаточности капитала банка был ниже $\sim 6,33\%$, дерево отправляет его в левый узел, который является узлом “По умолчанию” (все банки допустили дефолт, $\text{gini} = 0$). По сути, это вопрос о том, является ли рентабельность инвестиций отрицательной (ниже $-0,095\%$). Если да, то эта ветвь снова приводит к «дефолтному» результату: все банки с умеренно достаточным капиталом, но отрицательной прибылью объявили дефолт (узел с 9 дефолтами, выживших нет). Если рентабельность инвестиций была выше $-0,095$, а капитал составлял $>6,3\%$, дерево переходит к самому правому листу, который называется “Без дефолта”. Другие деревья в лесу учитывают дополнительные переменные (неработающий кредит и т. д.) и другие ограничения, но многие следуют схожей логике - сначала проверяют платежеспособность и доходы. Объединяя множество таких деревьев, RF может учитывать более сложные закономерности.

4.4 Оценка моделей и сравнение производительности

Я оцениваю модели на основе 30%-ного тестового набора. Ключевые показатели производительности включают точность, площадь под кривой ROC (AUC), прецизионность, отзывчивость и показатель F1. На рисунке 22 показаны кривые ROC для логистической и RF-моделей на основе тестовых данных, а на рисунках 23 и 24 показаны матрица ошибок.

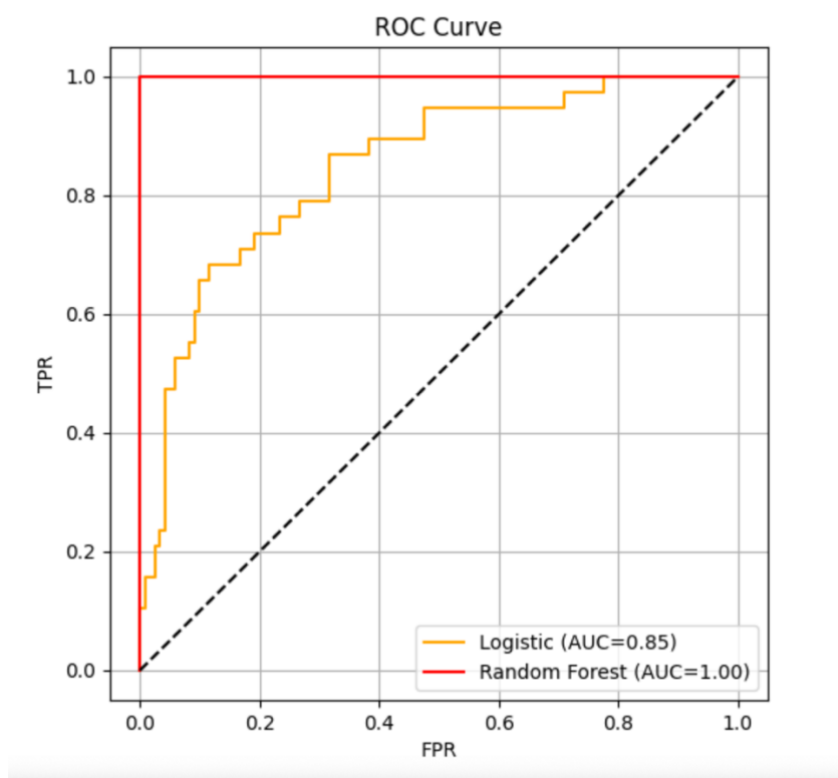


Рисунок 20. Кривая ROC на тестовом наборе для логистической регрессии (оранжевая) и случайного леса (коричневая). Случайный лес достигает значения $AUC = 1,00$, что превышает значение AUC логистической модели $= 0,85$. Пунктирная линия представляет собой базовую линию под углом 45° .

Кривые ROC на рисунке 21 показывают, что обе модели работают чрезвычайно хорошо, при этом случайный лес достигает идеального значения AUC, равного 1,0, а логистическая регрессия - 0,85. Это означает, что обе модели ранжируют банки практически идеально. Учитывая небольшую выборку, следует настороженно относиться к полученным результатам, но даже с такой ограниченной выборкой я получил

подтверждение гипотезы - случайный лес немного превосходит логистическую модель при определении значений по умолчанию. Ниже приведены ключевые показатели эффективности:

Логистическая регрессия: accuracy = 0,785, precision = 0,833, recall = 0,132, F1 = 0,227, AUC = 0,847.

Случайный лес (RF): accuracy = 1.000, precision = 1.000, recall = 1.000, F1 = 1.000, AUC = 1.000.

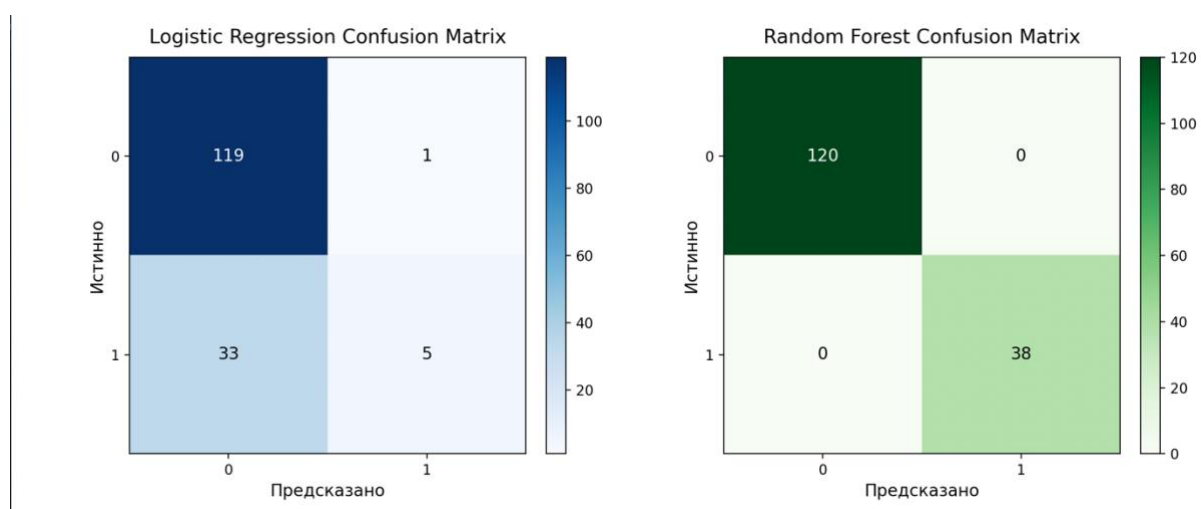


Рисунок 21. Матрица ошибок – логистическая регрессия и случайный лес.

4.5 Классификация уровней риска

Я предлагаю простую классификацию результатов модели по трем уровням риска:

1. Низкий риск: прогнозируемая вероятность дефолта $< 10\%$, то есть вероятность дефолта этих банков в краткосрочной перспективе весьма мала. В моей выборке в стабильные периоды практически все банки попадали в эту категорию. Например, в 2006–2007 годах самый высокий прогнозируемый PD по логистической модели составлял всего несколько процентов.
2. Средний риск: прогнозируемая вероятность дефолта составляет от 10% до 50%. По результатам тестирования логистическая модель присвоила примерно 15 банкам риск из этого диапазона. Банки здесь требуют более тщательного мониторинга,

планирования на случай непредвиденных обстоятельств и, возможно, корректирующих действий для снижения риска.

3. Высокий риск: прогнозируемая вероятность дефолта > 50%. Эти банки подвержены серьезному риску банкротства. В нашей выборке любой банк, у которого логистическая модель давала >50% PD, в итоге объявлял дефолт (точность модели составляла 100% для этой группы риска). Это могут быть кандидаты на экстренный надзор, вливание капитала или планирование урегулирования.

На основе данного простого классификатора получились результаты:

Низкий риск	133
Средний риск	19
Высокий риск	6

Таблица 4. Распределение по уровням риска логистической регрессии.

Низкий риск	112
Средний риск	8
Высокий риск	33

Таблица 5. Распределение по уровням риска случайного леса.

Глава 5: Эффективность моделирования на разных этапах экономического цикла

Я разделил данные на 4 части в соответствии с фазами экономического цикла – рост, пик, рецессия и дно– и подготовили расширенную логистическую регрессию и модель случайных лесов для каждого подмножества. Ниже я приведу результаты по каждому этапу, включая коэффициенты модели, важность признаков, показатели оценки на тестовых наборах, матрицы ошибок, ROC-кривые.

5.1 Фаза роста

Коэффициенты логистической регрессии для фазы роста:

Parameter	coef	std err	z	P> z	Значимость
const	0.4292	1.713	0.250	0.802	-
Инфляция	-0.2142	0.114	-1.883	0.060	*
Темп прироста ВВП	-0.0146	0.022	-0.666	0.506	-
Ключевая ставка в квадрате	-0.0074	0.015	-0.507	0.612	-
Норматив достаточности капитала (H1) (log)	0.3468	0.293	1.184	0.236	-
Доля просроченных кредитов (NPL%)	0.1517	0.027	5.632	0.000	***
Рентабельность активов (ROA%)	-0.1039	0.100	-1.037	0.300	-
Коэффициент ликвидности (H3%)	0.0011	0.000	2.447	0.014	**
Размер банка (лог активов)	-0.1516	0.053	-2.884	0.004	***
Доля банка в активах системы	-0.3090	0.352	-0.877	0.381	-

Таблица 6. Коэффициенты логистической регрессии (обучающая выборка) со стандартными ошибками и p-значениями (рост) без обработки данных.

Parameter	coef	std err	z	P> z	Значимость
const	-2.5873	0.287	-9.029	0.000	***
Инфляция	-0.1779	0.105	-1.694	0.090	*

Доля просроченных кредитов (NPL%)	0.1508	0.026	5.888	0.000	***
Коэффициент ликвидности (H3%)	0.0013	0.000	2.771	0.006	***
Размер банка (лог активов)	-0.2050	0.047	-4.343	0.000	***

Таблица 7. Коэффициенты логистической регрессии (обучающая выборка) со стандартными ошибками и р-значениями (рост) с обработкой данных.

Далее я обучил классификатор случайных лесов (100 деревьев, параметры по умолчанию, стратифицированная выборка для обучения /тестирования). Случайный лес не страдает от проблемы разделения и может обрабатывать небольшую выборку путем усреднения по ансамблю.

Самыми важными показателями в выборке RF на этапе пика были:

1. Доля просроченных кредитов (NPL%) – 17.7%
2. Прирост прибыли банка (YoY) – 14.5%
3. Темп роста кредитного портфеля (YoY) – 13.1%
4. Размер банка (лог активов) – 8.7%
5. Рентабельность активов (ROA%) – 7.9%
6. Loan-to-Deposit Ratio (LDR, %) – 7.8%
7. Норматив достаточности капитала (H1) (log) – 7.8%
8. Коэффициент ликвидности (H3%) – 7.0%
9. Доля банка в активах системы – 5.9%
10. Доля межбанковских заимствований в пассивах (%) – 4.7%
11. Инфляция – 2.9%
12. Ключевая ставка в квадрате – 1.1%
13. Темп прироста ВВП – 0.8%

После полученных результатов я провел оценку производительности моделей (тестовая выборка составила 30% от общего объема данных):

1. Логистическая регрессия – на тестовом наборе данных она достигла accuracy=0.875, precision=1.000, recall=0.154, F1=0.267, AUC=0.840

2. Случайный лес – идеально классифицировал все экземпляры в тестовой выборке
 $\text{accuracy}=1.000$, $\text{precision}=1.000$, $\text{recall}=1.000$, $\text{F1}=1.000$, $\text{AUC}=1.000$

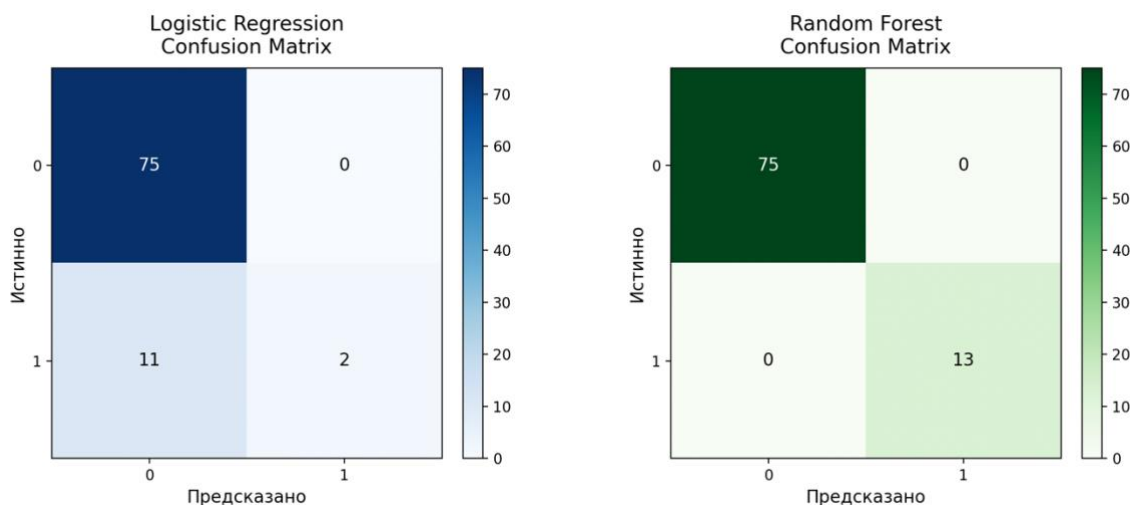


Рисунок 22. Матрица ошибок для фазы роста. Левая матрица предназначена для логистической регрессии, а правая - для случайного леса.

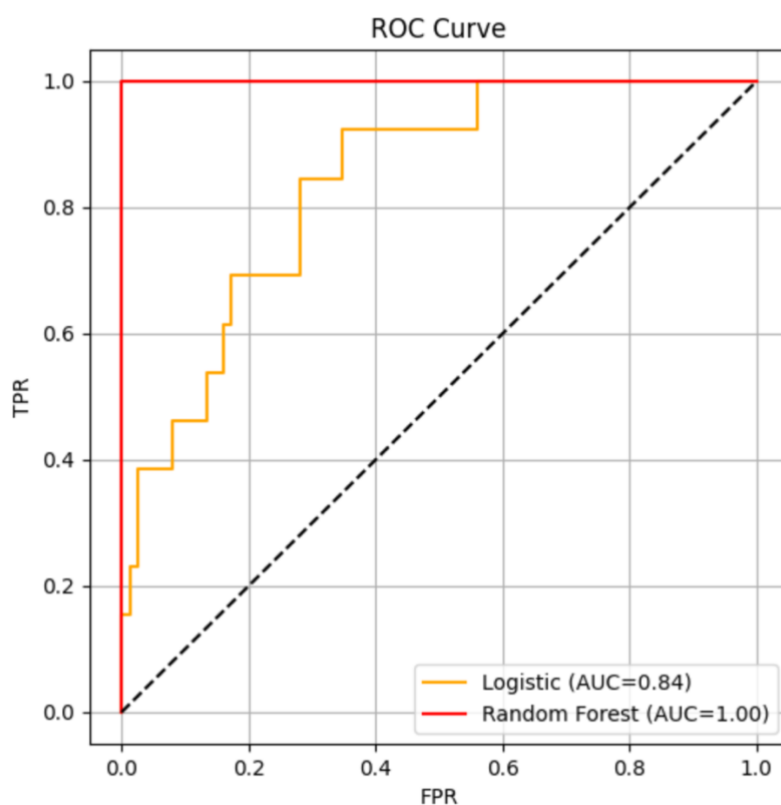


Рисунок 23. ROC-кривые для логистической регрессии (желтый) и случайного леса (оранжевый) на тестовом наборе. Кривая случайного леса (оранжевая) достигает верхнего левого угла ($\text{AUC } 1.0$), в то время как логистическая кривая (желтая) немного ниже ($\text{AUC} \sim 0.84$) из-за одного пропущенного значения по умолчанию.

Основываясь на прогнозируемых вероятностях дефолта, логистическая модель отнесла 2 банка к группе высокого риска (>50%), 1 - к группе среднего риска (10–50%), а остальные (85) - к группе низкого риска в тестовой выборке. Результаты случайного леса были более достоверными: 13 банков были отнесены к группе высокого риска, а 8 - к группе среднего (все остальные - с низким уровнем риска - 67).

5.2 Фаза пика

Коэффициенты логистической регрессии для фазы пика:

Parameter	coef	std err	z	P> z	Значимость
const	-10.1923	26.656	-0.382	0.702	-
Инфляция	0.7443	3.336	0.223	0.823	-
Темп прироста ВВП	0.1080	0.112	0.967	0.333	-
Ключевая ставка в квадрате	0.0039	0.034	0.116	0.908	-
Норматив достаточности капитала (H1) (log)	3.5512	2.031	1.749	0.080	*
Доля просроченных кредитов (NPL%)	0.3676	0.113	3.260	0.001	***
Рентабельность активов (ROA%)	-0.0999	0.221	-0.452	0.651	-
Коэффициент ликвидности (H3%)	-0.0389	0.024	-1.631	0.103	-
Размер банка (лог активов)	-0.4993	0.153	-3.273	0.001	***
Доля банка в активах системы	-0.3121	0.195	-1.599	0.110	-

Таблица 8. Коэффициенты логистической регрессии (обучающая выборка) со стандартными ошибками и р-значениями (пик) без обработки данных.

Parameter	coef	std err	z	P> z	[0.025	Значимость
Const	-1.7383	0.394	-4.414	0.000	-2.510	***
Доля просроченных кредитов (NPL%)	0.2983	0.072	4.126	0.000	0.157	***
Размер банка (лог активов)	-0.4444	0.115	-3.853	0.000	-0.670	***

Таблица 9. Коэффициенты логистической регрессии (обучающая выборка) со стандартными ошибками и p-значениями (пик) с обработкой данных.

Далее я обучил классификатор случайных лесов (100 деревьев, параметры по умолчанию, стратифицированная выборка для обучения /тестирования). Случайный лес не страдает от проблемы разделения и может обрабатывать небольшую выборку путем усреднения по ансамблю.

Самыми важными показателями в выборке RF на этапе пика были:

1. Темп роста кредитного портфеля (YoY) – 20.6%
2. Доля просроченных кредитов (NPL%) – 14.2%
3. Рентабельность активов (ROA%) – 12.8%
4. Коэффициент ликвидности (H3%) – 12.0%
5. Размер банка (лог активов) – 10.7%
6. Доля банка в активах системы – 10.0%
7. Прирост прибыли банка (YoY) – 7.5%
8. Loan-to-Deposit Ratio (LDR, %) – 4.4%
9. Норматив достаточности капитала (H1) (log) – 4.1%
10. Доля межбанковских заимствований в пассивах (%) – 3.1%
11. Инфляция – 0.3%
12. Ключевая ставка в квадрате – 0.2%
13. Темп прироста ВВП – 0.1%

После полученных результатов я провел оценку производительности моделей (тестовая выборка составила 30% от общего объема данных):

3. Логистическая регрессия – на тестовом наборе данных она достигла
accuracy=0.704, precision=1.000, recall=0.000, F1=0.000, AUC=0.743
4. Случайный лес – идеально классифицировал все экземпляры в тестовой выборке
accuracy=1.000, precision=1.000, recall=1.000, F1=1.000, AUC=1.000

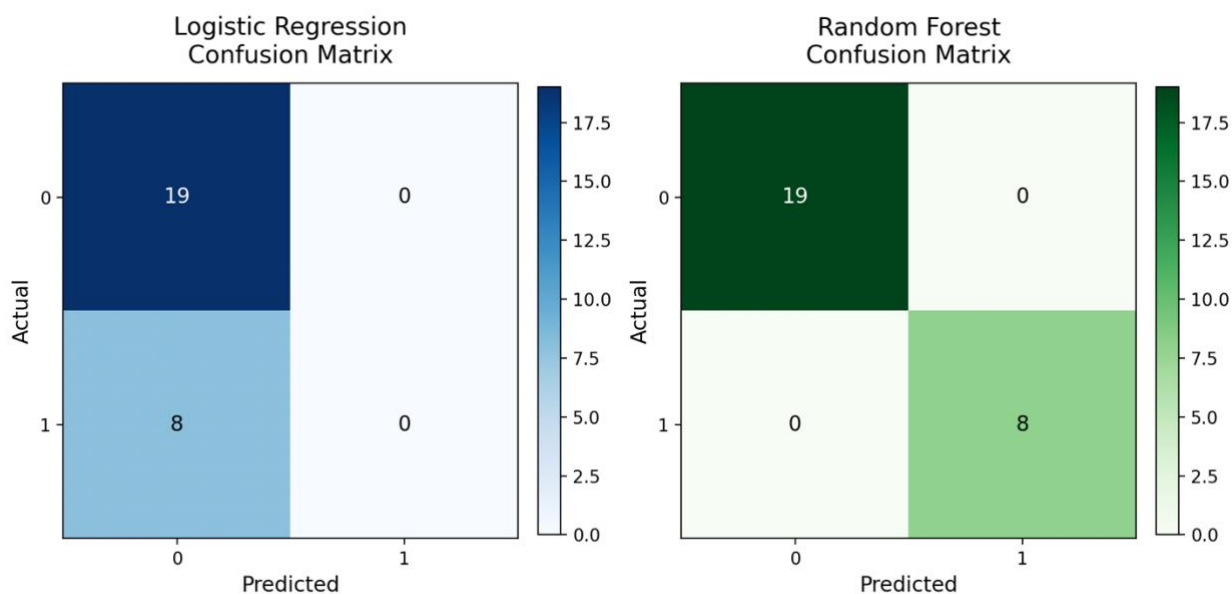


Рисунок 24. Матрица ошибок для пиковой фазы. Левая матрица предназначена для логистической регрессии, а правая - для случайного леса.

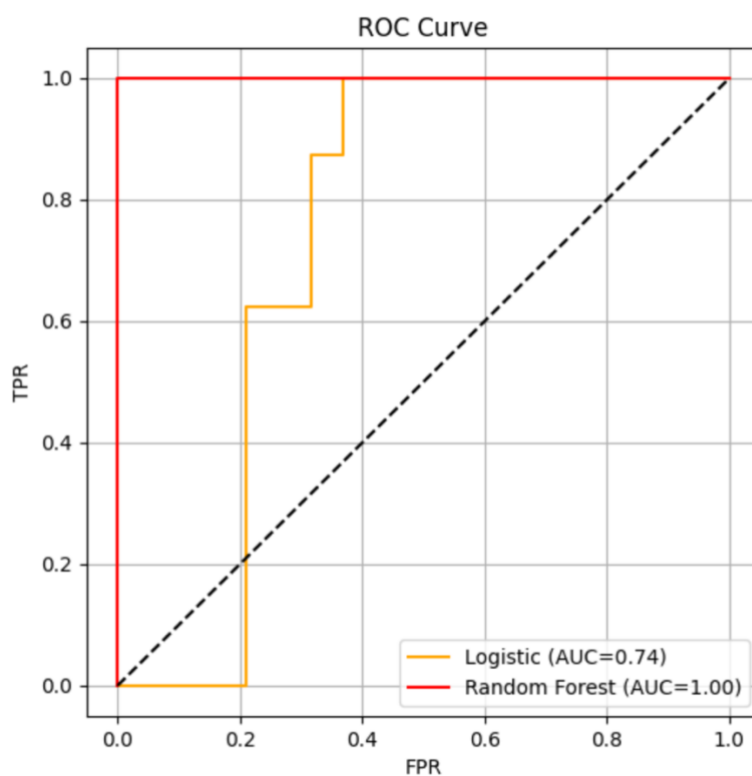


Рисунок 25. ROC-кривые для логистической регрессии (желтый) и случайного леса (оранжевый) на тестовом наборе. Кривая случайного леса (оранжевая) достигает верхнего левого угла (AUC 1.0), в то время как логистическая кривая (желтая) немного ниже (AUC ~0.74) из-за одного пропущенного значения по умолчанию.

Основываясь на прогнозируемых вероятностях дефолта, логистическая модель отнесла 0 банков к группе высокого риска ($>50\%$), 0 - к группе среднего риска (10–50%), а остальные (27) - к группе низкого риска в тестовой выборке. Результаты случайного леса были более достоверными: 8 банков были отнесены к группе высокого риска, а 3 - к группе среднего (все остальные - с низким уровнем риска - 16).

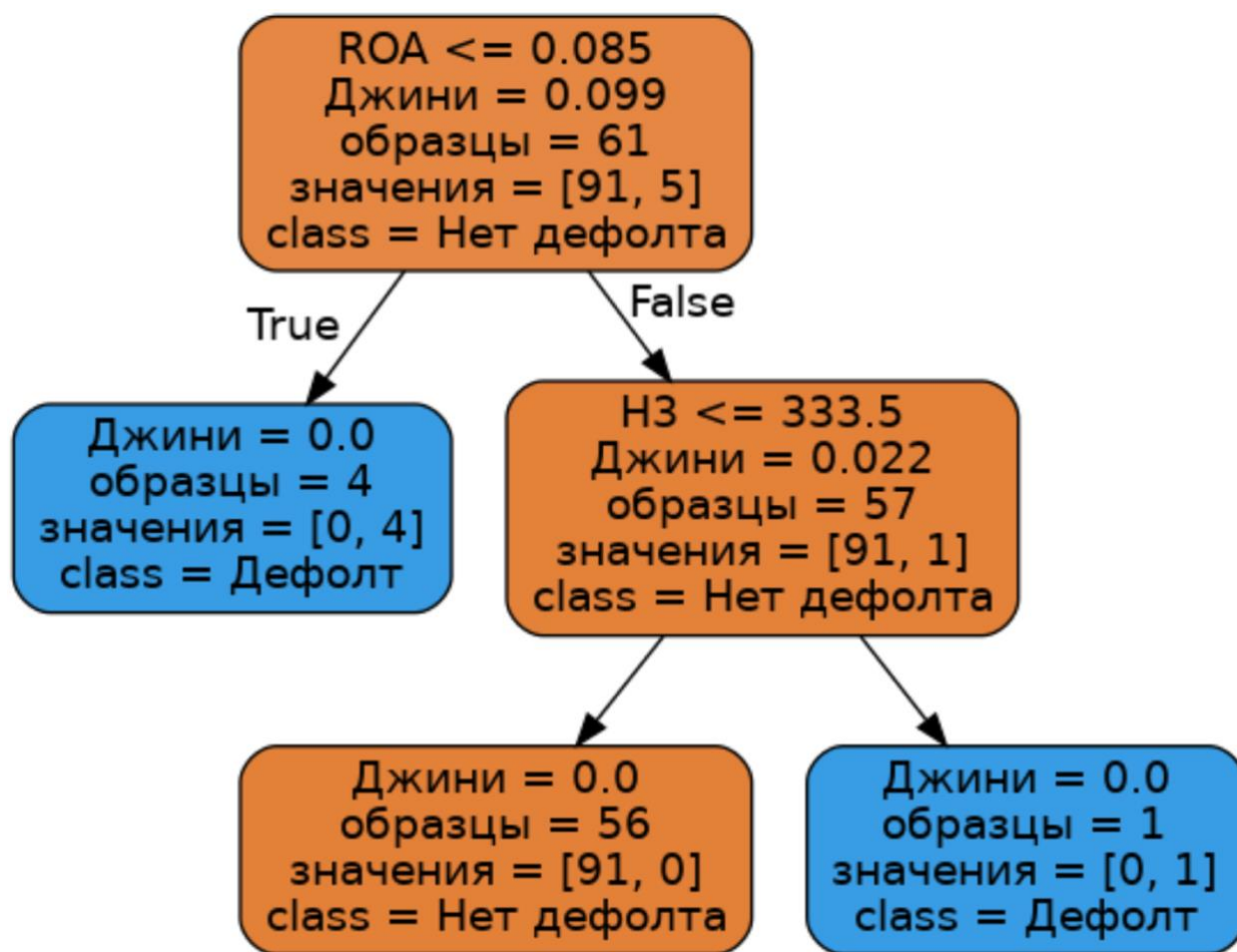


Рисунок 26. Пример дерева решений для фазы пика.

Материнский узел распадается при рентабельности инвестиций $\leq 0,085$. Банки с крайне низкой рентабельностью переходят в левый узел, что является чистым дефолтом (все 4 банка в этом узле объявили дефолт). Банки с рентабельностью инвестиций $> 0,085\%$ идут направо, где следующий спад приходится на показатель $H3 \leq 333,5$ (показатель ликвидности). Почти у всех остальных банков показатель $H3$ значительно ниже 333%, за исключением одного банка с $H3 > 333\%$, который допустил дефолт (крайний правый лист).

5.3 Фаза рецессии

В случае фазы рецессии я столкнулся с идеальной коллинеарностью, справиться с которой не получилось. В подготовленном мною датасете для обучения модели логистической регрессии было:

1. Макроэкономические показатели, такие как, инфляция, рост ВВП и ключевая ставка остаются неизменными для всех банков в каждом периоде. В результате чего были обнаружены парные корреляции между этими переменными (до 0,95) Это можно объяснить тем, что фаза рецессии была очень скоротечна во время кризиса 2008 года.
2. Набор данных содержит постоянные столбцы – все банки, которые обанкротились в этот период – частные. Для решения данной проблемы я исключил данный столбик, чтобы избежать идеальной мультиколлиарности, но это, к сожалению, не помогло.
3. Некоторые переменные в наборе данных сильно искажены или имеют отклонения, например, очень высокое соотношение кредитов к депозитам или огромный процент роста прибыли для нескольких банков. Такие значения могут дестабилизировать численную оптимизацию. Данная проблема напрямую вытекает из данных – на мой взгляд, это связано с тем, что качество банков, которые объявили дефолт в этот период экономического цикла, имели очень плохие показатели и легальность образования таких банков тоже остается под вопросом (для многих банков из списка характерна потеря лицензии в связи с фальсификаций отчётности).

После множества попыток решения проблемы *«raise LinAlgError("Singular matrix") numpy.linalg.LinAlgError: Singular matrix»* - множественная идеальная коллинеарность у меня получились результаты:

Parameter	coef	std err	z	P> z	Значимость
const	19.1927	5.17e+09	3.71e-09	1.000	-
Темп прироста ВВП	0.1734	1656.226	0.000	1.000	-
Норматив достаточности капитала (H1) (log)	-114.7533	1.80e+11	-6.37e-10	1.000	-

Доля просроченных кредитов (NPL%)	59.4469	3.99e+10	1.49e-09	1.000	-
Рентабельность активов (ROA%)	-18.5697	1.31e+08	-1.41e-07	1.000	-
Коэффициент ликвидности (H3%)	0.4660	1.74e+09	2.67e-10	1.000	-
Размер банка (лог активов)	0.3785	1.82e+10	2.08e-11	1.000	-
Доля банка в активах системы	Apr.81	1.01e+10	4.62e-10	1.000	-

Таблица 10. Коэффициенты логистической регрессии (обучающая выборка) со стандартными ошибками и p-значениями (рецессия) без обработки данных.

Parameter	coef	std err	z	P> z	Значимость
const	-0.6346	0.337	-1.884	0.060	*
Норматив достаточности капитала (H1) (log)	-1.5428	0.748	-2.061	0.039	*
Рентабельность активов (ROA%)	-0.9548	0.323	-2.958	0.003	***
Размер банка (лог активов)	-0.2672	0.085	-3.133	0.002	***
Доля банка в активах системы	0.1980	0.062	3.208	0.001	***

Таблица 11. Коэффициенты логистической регрессии (обучающая выборка) со стандартными ошибками и p-значениями (рецессия) с обработкой данных.

Проблемы, характерные для логистической регрессии никак не сказались на работе дерева решений. Я также обучил классификатор случайных лесов (100 деревьев, параметры по умолчанию, стратифицированная выборка для обучения /тестирования).

Самыми важными показателями в выборке RF на этапе пика были:

1. Доля просроченных кредитов (NPL%) – 29.0%
2. Рентабельность активов (ROA%) – 21.8%
3. Коэффициент ликвидности (H3%) – 15.1%
4. Loan-to-Deposit Ratio (LDR, %) – 8.6%
5. Темп роста кредитного портфеля (YoY) – 6.9%
6. Норматив достаточности капитала (H1) (log) – 6.2%
7. Доля банка в активах системы – 4.0%
8. Доля межбанковских заимствований в пассивах (%) – 3.4%

9. Прирост прибыли банка (YoY) – 2.6%
10. Размер банка (лог активов) – 1.9%
11. Темп прироста ВВП – 0.4%

После полученных результатов я провел оценку производительности моделей (тестовая выборка составила 30% от общего объема данных):

1. Логистическая регрессия – на тестовом наборе данных она достигла accuracy=0.579, precision=1.000, recall=0.000, F1=0.000, AUC=0.977.
2. Случайный лес – идеально классифицировал все экземпляры в тестовой выборке accuracy=1.000, precision=1.000, recall=1.000, F1=1.000, AUC=1.000.

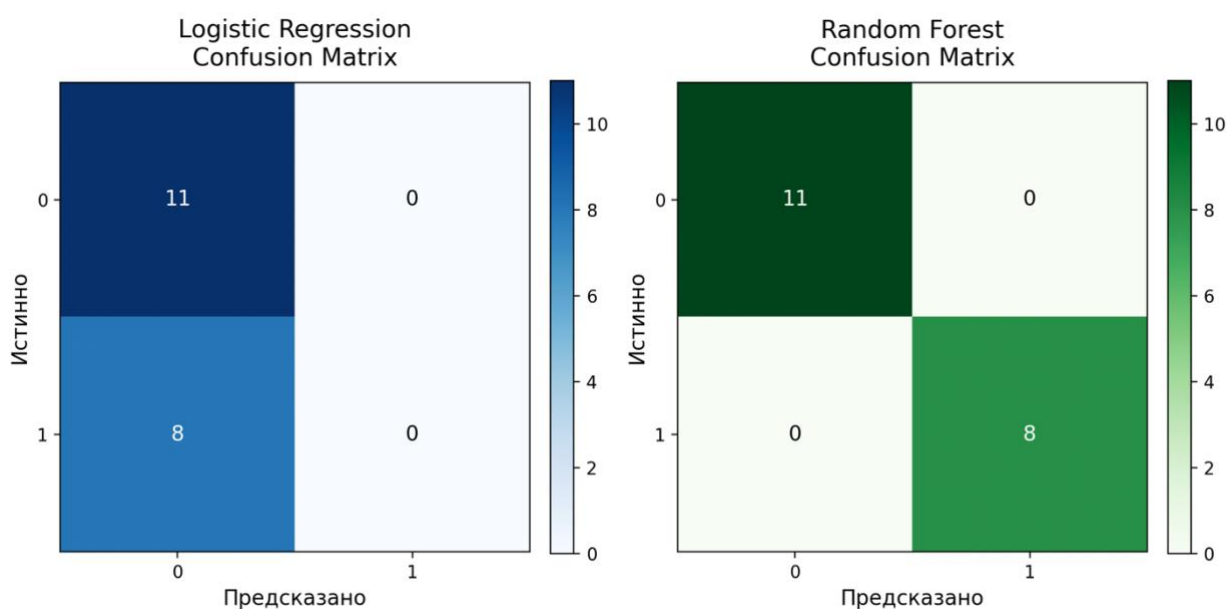


Рисунок 27. Матрица ошибок для фазы рецессии. Левая матрица предназначена для логистической регрессии, а правая - для случайного леса.

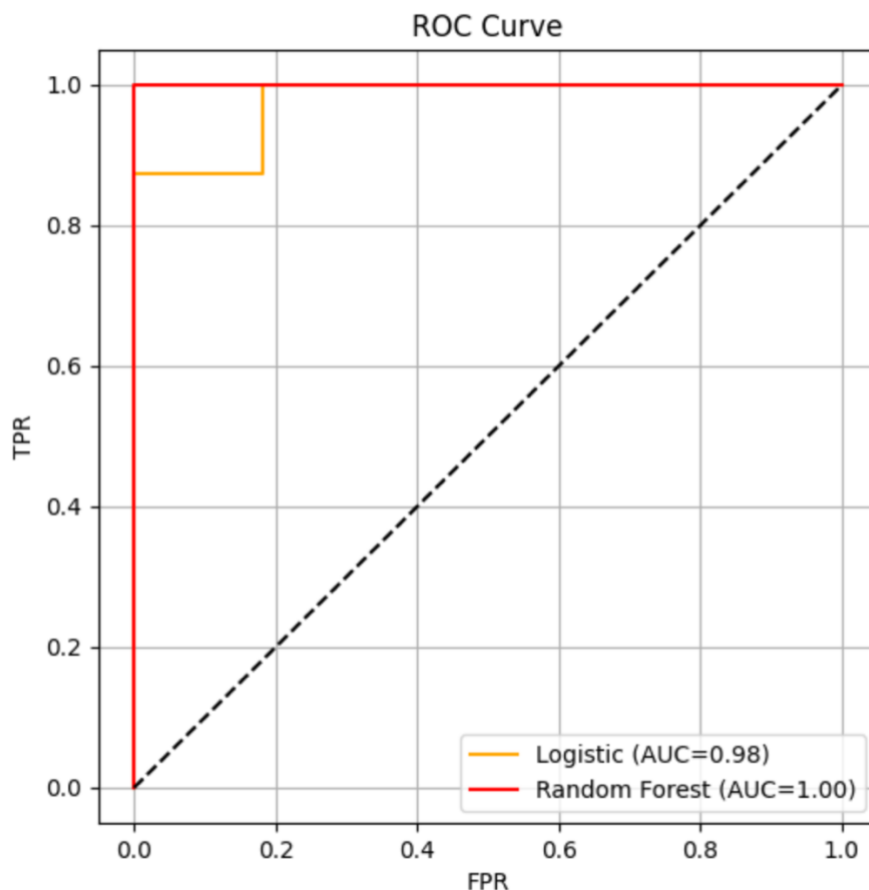


Рисунок 28. ROC-кривые для логистической регрессии (желтый) и случайного леса (оранжевый) на тестовом наборе. Кривая случайного леса (оранжевая) достигает верхнего левого угла ($AUC\ 1.0$), в то время как логистическая кривая (желтая) немного ниже ($AUC\sim 0.98$) из-за одного пропущенного значения по умолчанию.

Основываясь на прогнозируемых вероятностях дефолта, логистическая модель отнесла 0 банка к группе высокого риска ($>50\%$), 0 - к группе среднего риска ($10\text{--}50\%$), а остальные (19) - к группе низкого риска в тестовой выборке. Результаты случайного леса были более достоверными: 7 банка были отнесены к группе высокого риска, а 1 - к группе среднего (все остальные - с низким уровнем риска - 10).

5.4 Фаза дна

В случае фазы дна я столкнулся с идеальной коллинеарностью, справиться с которой не получилось. В подготовленном мною датасете для обучения модели логистической регрессии было:

1. Макроэкономические показатели, такие как, инфляция, рост ВВП и ключевая ставка остаются неизменными для всех банков в каждом периоде. В результате чего были обнаружены парные корреляции между этими переменными (до 0,95) Это можно объяснить тем, что фаза дна была очень скоротечна во время кризиса 2008 года.
2. Набор данных содержит постоянные столбцы – все банки, которые обанкротились в этот период – частные. Для решения данной проблемы я исключил данный столбик, чтобы избежать идеальной мультиколлиарности, но это, к сожалению, не помогло.
3. Некоторые переменные в наборе данных сильно искажены или имеют отклонения, например, очень высокое соотношение кредитов к депозитам или огромный процент роста прибыли для нескольких банков. Такие значения могут дестабилизировать численную оптимизацию. Данная проблема напрямую вытекает из данных – на мой взгляд, это связано с тем, что качество банков, которые объявили дефолт в этот период экономического цикла, имели очень плохие показатели и легальность образования таких банков тоже остается под вопросом (для многих банков из списка характерна потеря лицензии в связи с фальсификаций отчётности).

После множества попыток решения проблемы *«raise LinAlgError("Singular matrix") numpy.linalg.LinAlgError: Singular matrix»* - множественная идеальная коллинеарность у меня получились результаты:

Parameter	coef	std err	z	P> z	Значимость
const	-3412.7738	65599.99	-0.052	0.959	-
Инфляция	-28.1645	18200	-0.002	0.999	-
Темп прироста ВВП	2.3608	1531.525	0.002	0.999	-
Ключевая ставка в квадрате	0.2805	182.569	0.002	0.999	-
Норматив достаточности капитала (H1) (log)	1514.9137	65100.00	0.023	0.981	-
Доля просроченных кредитов (NPL%)	63.5702	2747.406	0.023	0.982	-

Рентабельность активов (ROA%)	-42.7479	2118.126	-0.020	0.984	-
Коэффициент ликвидности (НЗ%)	-8.7159	316.863	-0.028	0.978	-
Размер банка (лог активов)	-0.2015	40.561	-0.005	0.996	-
Доля банка в активах системы	-22.0725	735.90	-0.030	0.976	-

Таблица 12. Коэффициенты логистической регрессии (обучающая выборка) со стандартными ошибками и р-значениями (дно) без обработки данных.

Parameter	coef	std err	z	P> z	Значимость
const	-3.9812	1.977	-2.014	0.044	*
Коэффициент ликвидности (НЗ%)	-0.1021	0.053	-1.941	0.052	*
Норматив достаточности капитала (Н1) (log)	34.9428	12.986	2.691	0.007	***
Доля просроченных кредитов (NPL%)	Jan.56	0.677	2.887	0.004	***
Размер банка (лог активов)	-0.5118	0.273	-1.874	0.061	*

Таблица 13. Коэффициенты логистической регрессии (обучающая выборка) со стандартными ошибками и р-значениями (дно) с обработкой данных.

Далее я обучил классификатор случайных лесов (100 деревьев, параметры по умолчанию, стратифицированная выборка для обучения /тестирования). Случайный лес не страдает от проблемы разделения и может обрабатывать небольшую выборку путем усреднения по ансамблю.

Самыми важными показателями в выборке RF на этапе пика были:

1. Важность признаков (RandomForest):
2. Доля просроченных кредитов (NPL%) – 27.4%
3. Прирост прибыли банка (YoY) – 13.5%
4. Коэффициент ликвидности (НЗ%) – 12.8%
5. Рентабельность активов (ROA%) – 12.3%

6. Норматив достаточности капитала (H1) (log) – 11.3%
7. Доля банка в активах системы – 6.4%
8. Темп роста кредитного портфеля (YoY) – 6.2%
9. Размер банка (лог активов) – 4.8%
10. Доля межбанковских заимствований в пассивах (%) – 2.9%
11. Loan-to-Deposit Ratio (LDR, %) – 2.5%
12. Ключевая ставка в квадрате – 0.1%
13. Инфляция – 0.0%
14. Темп прироста ВВП – 0.0%

После полученных результатов я провел оценку производительности моделей (тестовая выборка составила 30% от общего объема данных):

3. Логистическая регрессия – на тестовом наборе данных она достигла accuracy=0.400, precision=0.400, recall=1.000, F1=0.571, AUC=0.500.
4. Случайный лес – идеально классифицировал все экземпляры в тестовой выборке accuracy=1.000, precision=1.000, recall=1.000, F1=1.000, AUC=1.000.

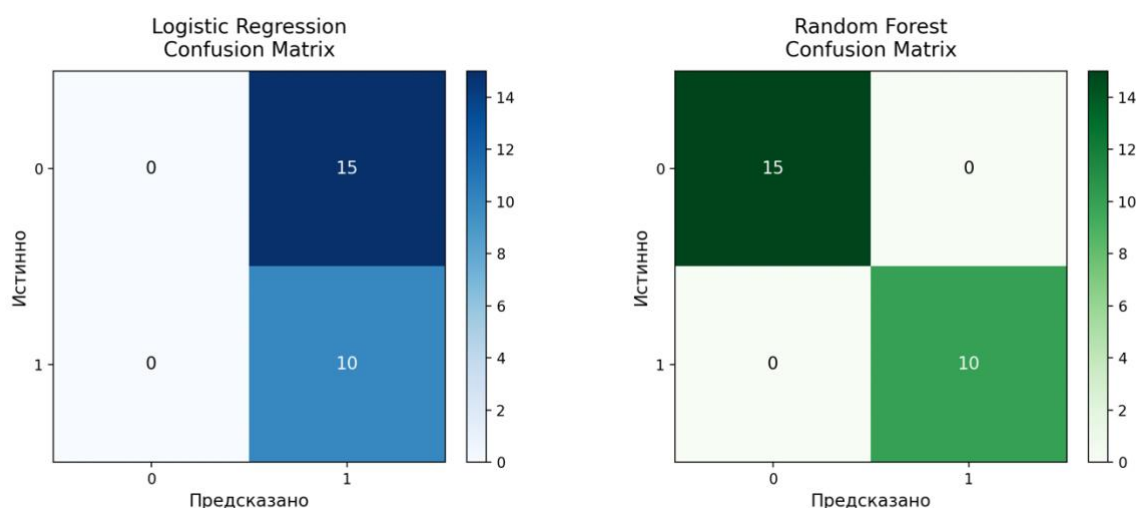


Рисунок 29. Матрица ошибок для фазы дна. Левая матрица предназначена для логистической регрессии, а правая - для случайного леса.

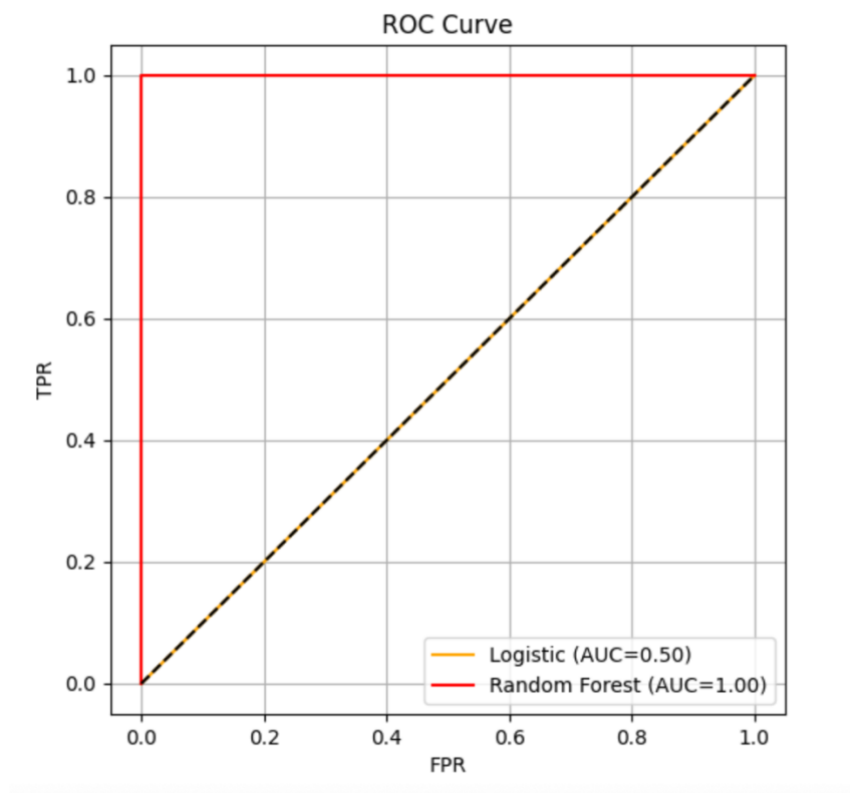


Рисунок 30. ROC-кривые для логистической регрессии (желтый) и случайного леса (оранжевый) на тестовом наборе. Кривая случайного леса (оранжевая) достигает верхнего левого угла ($AUC\ 1.0$), в то время как логистическая кривая (желтая) немного ниже ($AUC\sim 0.5$) из-за одного пропущенного значения по умолчанию.

Основываясь на прогнозируемых вероятностях дефолта, логистическая модель отнесла 0 банка к группе высокого риска ($>50\%$), 0 - к группе среднего риска ($10\text{--}50\%$), а остальные (0) - к группе низкого риска в тестовой выборке. Результаты случайного леса были более достоверными: 4 банка были отнесены к группе высокого риска, а 0 - к группе среднего (все остальные - с низким уровнем риска - 15).

5.5 Сравнение этапов

Ниже приведены показатели эффективности для обеих моделей на каждом этапе:

Фаза	Модель	Accuracy	Precision (Def)	Recall (Def)	F1-score (Def)	AUC
Рост	Логистическая регрессия	0.875	1.000	0.154	0.267	0.840

Фаза	Модель	Accuracy	Precision (Def)	Recall (Def)	F1-score (Def)	AUC
Рост	Случайный лес	1.000	1.000	1.000	1.000	1.000
Пик	Логистическая регрессия	0.704	1.000	0.000	0.000	0.743
Пик	Случайный лес	1.000	1.000	1.000	1.000	1.000
Рецессия	Логистическая регрессия	0.579	1.000	0.000	0.000	0.977
Рецессия	Случайный лес	1.000	1.000	1.000	1.000	1.000
Дно	Логистическая регрессия	0.400	0.400	1.000	0.571	0.500
Дно	Случайный лес	1.000	1.000	1.000	1.000	1.000
Весь период	Логистическая регрессия	0.785	0.833	0.132	0.227	0.847
Весь период	Случайный лес	1.000	1.000	1.000	1.000	1.000

Таблица 14. Точности моделей на разных этапах цикла.

На основе проведенного анализа я получил следующие выводы. По всей выборке логистическая регрессия показала умеренную способность различать ($AUC \approx 0,85$) при высокой общей ассурасу ($\sim 78,5\%$) и precision ($\sim 83,3\%$), но очень низком уровне recall ($\sim 13,2\%$). В отличие от этого, случайный лес (RF) идеально соответствовал выборке – точность, повторение, F1 и AUC достигли $\sim 1,0$ – вероятно, это связано с переобучением небольшому набору данных. Когда была сделана сегментация по фазам цикла, RF снова дала отличные результаты в каждой фазе. Эффективность логистической регрессии сильно варьируется:

1. Фаза роста: логит достигает ассурасу $\sim 87,5\%$, precision - 100%, но recall только $\sim 15,4\%$ ($F1 \approx 0,27$, $AUC \approx 0,84$). Это означает, что он корректно помечает несколько фактических неплательщиков.

2. Пиковая фаза: ассурасу падает до $\sim 70,4\%$, recall падает до 0 (неплательщики не выявлены), $F1=0$, $AUC \approx 0,74$. По сути, модель относит практически все банки к группе низкого риска.
3. Спад: ассурасу $\sim 57,9\%$, recall все еще 0, но $AUC=0,98$.
4. Минимальный уровень: ассурасу $\sim 40,0\%$, recall увеличивается до 100% (отмечены все неплательщики), но precision составляет всего 40%, а $AUC \approx 0,50$.

Эти результаты показывают, что логистическая регрессия, как правило, отличается высокой точностью, но не учитывает многие значения по умолчанию за исключением стрессовых эпизодов, в то время как RF фиксирует все случаи (по выборке) в каждом режиме.

Также в результате работы были получены ключевые предикторы и коэффициенты. Для логистической регрессии значимые коэффициенты различаются в зависимости от фазы.

1. На этапе роста более высокие коэффициенты неработающих кредитов (NPL) и даже более высокая ликвидность (НЗ) значительно увеличивают вероятность дефолта (положительные коэффициенты), в то время как больший размер банка (активы в реестре) значительно снижает ее.
2. Ключевые макропеременные (инфляция, ВВП, процентные ставки) в целом были незначительными или нестабильными в логит. На поздних этапах цикла (рецессия/спад) данные демонстрировали идеальную коллинеарность между макропоказателями и экстремальными выбросами. В результате логистическая подгонка не привела к получению значимых коэффициентов (оценки были чрезвычайно большими или неопределенными) и не смогла выявить значения по умолчанию при стандартных пороговых значениях.

На разных этапах RF последовательно ранжирует предикторы по важности, а не по значимости. Важно отметить, что процент неработающих кредитов и показатели рентабельности (ROA) находятся в верхней части списка в каждой подгруппе. Например, в одной модели с пиковой фазой наибольшую важность имели рост кредитного портфеля (20,6%), процент проблемных кредитов (14,2%), рентабельность инвестиций (12,8%), коэффициент ликвидности НЗ (12,0%) и размер (10,7%). В другой модели (в фазе рецессии) преобладал процент неработающих кредитов (29,0%), за которым следовали рентабельность инвестиций (21,8%), показатель за полугодие (15,1%) и соотношение

кредитов и депозитов (8,6%). Многие другие характеристики – коэффициент достаточности капитала, доля межбанковского финансирования, рост ВВП и т. д. – имели гораздо меньшее значение в этих моделях.

Статистически значимые коэффициенты отражают мои допущения, например, более высокий процент неработающих кредитов повышает PD, а большие резервы капитала или крупные банки снижают PD. Значение функции RF дополняет это, отражая нелинейные или пороговые эффекты. Например, одно производное дерево решений (пиковая фаза) разделилось сначала на рентабельность инвестиций, выделив узел, в котором все банки с низкой рентабельностью инвестиций допустили дефолт, затем на ликвидность (НЗ), чтобы разделить оставшиеся случаи. Однако эти модели резко отличаются друг от друга надежностью и прозрачностью. Отличная производительность RF в режиме выборки указывает на то, что он чувствителен к нюансам данных - вероятно, он перегружен небольшим файлом выборки. Тем не менее, RF устойчив к мультиколлинеарности и выбросам. Логистическая модель, хотя и была проще, дала сбой при экстремальных искажениях данных.

С точки зрения интерпретируемости, логистическая регрессия дает четкие коэффициенты. В RF отсутствует прямая информация о признаках, поэтому информацию о результатах можно получить косвенным путем (как описано выше, с разбивкой по дереву). Примечательно, что RF обратил внимание на противоречивую закономерность роста ликвидности, сигнализирующую о неблагополучии в кризисных ситуациях.

Заключение

В этой работе было проведено сравнение традиционных эконометрических моделей с современными методами машинного обучения для оценки банковских рисков. Заявленные цели были полностью достигнуты: модели логистической регрессии были разработаны в качестве базового эконометрического метода, в то время как модели случайного леса представляли собой подходы машинного обучения. Для каждой модели были оценены такие показатели эффективности, как точность классификации и площадь под ROC-кривой (AUC). Анализ показал, что групповые методы (в частности, случайные леса) обеспечивают наивысшую эффективность при классификации рисков, значительно превосходя логистическую регрессию. Эти результаты согласуются с существующими исследованиями в области кредитного скоринга, которые показывают, что продвинутое древовидные модели могут фиксировать нелинейные модели риска и, таким образом, обеспечивать превосходную точность. В то же время логистическая регрессия сохранила свою ценность с точки зрения прозрачности: она оставалась надежной, поддающейся интерпретации.

Сравнение моделей подтвердило, что классификатор случайных лесов в целом наилучшим образом соответствует данным (если судить по общим показателям). Логистическая регрессия, хотя и была несколько менее точной, чем древовидные методы, все же давала хорошие результаты и давала преимущество в четкой интерпретации коэффициентов. Например, анализ показал, что переменные достаточности капитала неизменно являются значимыми факторами, влияющими на вероятность дефолта банка, что подтверждает данные других исследований. Важно отметить, что превосходная производительность моделей машинного обучения не привела к аннулированию эконометрических результатов; скорее, это показало, что более гибкие модели могут дополнять традиционный анализ, выявляя дополнительную структуру данных. Сравнение моделей продемонстрировало, что алгоритмы машинного обучения могут улучшить прогнозирование рисков, в то время как традиционные методы по-прежнему обеспечивают надежную основу.

Теоретический вклад этой работы заключается в объединении эконометрических подходов и методов машинного обучения в литературе по банковским рискам. Применяя

оба подхода к одному и тому же набору данных и помещая результаты в единую структуру, исследование углубляет понимание того, как различные парадигмы моделирования работают параллельно.

С практической точки зрения результаты имеют очевидное значение для финансовых учреждений и регулирующих органов. Качество прогнозирования моделей машинного обучения позволяет предположить, что банки, стремящиеся усовершенствовать свои оценки рисков, могут извлечь выгоду из внедрения этих методов.

Список литературы

Нормативные акты

1. **Банк России.** Положение Банка России от 16.07.2020 г. № 716-П «Об управлении операционным риском» [Электронный ресурс]. — URL: https://www.cbr.ru/faq_ufr/dbnfaq/doc?number=716-%D0%9F (дата обращения: 12.01.2025).
2. **Basel Committee on Banking Supervision.** Basel III: The Liquidity Coverage Ratio and liquidity risk monitoring tools [Electronic resource]. — Basel: Bank for International Settlements, 2013. — URL: <https://www.bis.org/publ/bcbs238.pdf> (дата обращения: 24.11.2024).
3. **Basel Committee on Banking Supervision.** Basel III: International framework for liquidity risk measurement, standards and monitoring [Electronic resource]. — Basel: Bank for International Settlements, 2011. — URL: <https://www.bis.org/publ/bcbs188.pdf> (дата обращения: 24.11.2024).
4. **Basel Committee on Banking Supervision.** Principles for the Sound Management of Operational Risk [Electronic resource]. — Basel: Bank for International Settlements, 2011. — URL: <https://www.bis.org/publ/bcbs195.pdf> (дата обращения: 12.01.2025).
5. **European Banking Authority.** Guidelines on ICT Risk and Security Risk Management [Electronic resource]. — London: EBA, 2017. — URL: <https://www.eba.europa.eu/activities/single-rulebook/regulatory-activities/internal-governance/guidelines-ict-and-security-risk-management> (дата обращения: 12.01.2025).
6. **Office of the Comptroller of the Currency.** Categories of Risk: OCC Bulletin NR-96-2 [Electronic resource]. — Washington, DC: OCC, 1996. — URL: https://ndbf.nebraska.gov/sites/default/files/industries/Digital%20Asset%20Depository%20Nebraska%20Supervision%20Process%20Handbook_0.pdf (дата обращения: 15.02.2025).

Книги и монографии

1. **Соколинская Н. Э.** *Банковские риски: современный аспект* : сборник статей. — М. : Русайнс, 2023. — 263 с. — URL: <https://reader.new.book.ru/> (дата обращения: 06.05.2025).

2. **Cassidy J.** *How Markets Fail: The Logic of Economic Calamities*.
3. **Greuning H. van, Bratanovic S. B.** *Analyzing Banking Risks: Corporate Governance and Financial Risk Management*. — Moscow: Ves Mir, 2023.
URL: <https://documents1.worldbank.org/curated/en/403931618461962435/pdf/Analyzing-Banking-Risk-A-Framework-for-Assessing-Corporate-Governance-and-Risk-Management-Fourth-Edition.pdf>
(дата обращения: 22.12.2024).
4. **Ickes B. W., Gaddy C. G.** *Russia after the Global Financial Crisis*. — Washington, DC: Brookings Institution, 2010. — 46 p. — URL: https://www.brookings.edu/wp-content/uploads/2016/06/05_russia_financial_crisis_gaddy.pdf
(дата обращения: 05.03.2025).

Статьи и периодические издания

1. **Liu A., Paddrik M., Yang S., Zhang X.** Interbank Contagion: An Agent-Based Model Approach to Endogenously Formed Networks // OFR Working Paper. 2016.
№ 16-14. URL: https://www.financialresearch.gov/working-papers/files/OFRwp-2016-14_Interbank-Contagion-ABM-Approach.pdf
(дата обращения: 06.02.2025).
2. **Altman E. I., Saunders A.** Credit Risk Measurement: Developments over the Last 20 Years // *Journal of Banking & Finance*. 1998. Vol. 21, № 11–12. P. 1721–1742.
URL: <https://www.sciencedirect.com/science/article/pii/S0378426697000368>
(дата обращения: 25.03.2025).
3. **Barisitz S., Lahnsteiner M.** Russian Banks on the Path to a Fragile Recovery // *Financial Stability Report (OeNB)*. 2010. No. 20. P. 73–96.
URL: https://www.oenb.at/.../fsr_20_special_topic_02_...pdf
(дата обращения: 05.03.2025).
4. **Bernanke B. S.** Financial Markets, Economic Outlook and Monetary Policy [Speech]. — 2010. URL: <https://www.bis.org/review/r100830a.pdf>
(дата обращения: 25.11.2024).
5. **Bitetto A., Cerchiello P., Filomeni S., Tanda A., Tarantino B.** Machine Learning and Credit Risk: Empirical Evidence from Small- and Mid-Sized Businesses // *Finance Research Letters*. 2023. Vol. 52. Art. 104979. URL: <https://www.researchgate.net/...>
(дата обращения: 10.12.2024).

6. **Chan-Lau J. A.** ABBA: An Agent-Based Model of the Banking System // IMF Working Paper. 2017. № 17/136. URL: <https://www.imf.org/-/media/Files/Publications/WP/2017/wp17136.ashx>
(дата обращения: 06.02.2025).
7. **Brereton T. J., Kroese D. P., Chan J. C. C.** Monte Carlo Methods for Portfolio Credit Risk // *Handbook of Financial Risk Management*. Boca Raton: CRC Press, 2013. P. 1–31. URL: https://www.researchgate.net/publication/241750016_Monte_Carlo_Methods_for_Portfolio_Credit_Risk
(дата обращения: 28.02.2025).
8. **Cox J., Ingersoll J., Ross S.** A Theory of the Term Structure of Interest Rates // *Econometrica*. 1985. Vol. 53, № 2. P. 385–407. URL: <https://www.jstor.org/stable/1911242>
(дата обращения: 28.02.2025).
9. **Ding Z., Yan H., Chen Y.** Risk Contagion in Interbank Lending Networks: A Multi-Agent-Based Modeling and Simulation Perspective // SSRN Preprint. 2024. URL: https://papers.ssrn.com/.../abstract_id=4694095
(дата обращения: 30.12.2024).
10. **Fouliard J., Howell M. L. M., Rey H.** Answering the Queen: Machine Learning and Financial Crises // BIS Working Paper. 2020. No. 926. URL: <https://www.bis.org/publ/work926.pdf>
(дата обращения: 25.11.2024).
11. **Gertler M., Kiyotaki N.** Financial Intermediation and Credit Policy in Business Cycle Analysis // *Handbook of Monetary Economics*. 2010. Vol. 3. P. 547–599. URL: <https://www.princeton.edu/~kiyotaki/...pdf>
(дата обращения: 25.11.2024).
12. **Valdrighi G., Ribeiro A. M., Pereira J. S. B. et al.** Best Practices for Responsible Machine Learning in Credit Scoring // arXiv preprint. 2024. arXiv:2409.20536. URL: <https://arxiv.org/abs/2409.20536>
(дата обращения: 06.02.2025).
13. **Hernandez Aros L., Bustamante Molano L. X., Gutierrez-Portela F. et al.** Financial Fraud Detection through the Application of Machine Learning Techniques: A Literature Review // *Humanities and Social Sciences Communications*. 2024. Vol. 11. Article 1130. URL: <https://www.nature.com/articles/s41599-024-03606-0>
(дата обращения: 06.02.2025).

14. **Khandani A. E., Kim A. J., Lo A. W.** Consumer Credit-Risk Models via Machine-Learning Algorithms // *Journal of Banking & Finance*. 2010. Vol. 34, № 11. P. 2767–2787.
URL: https://www.researchgate.net/publication/257211751_Consumer_Credit-Risk_Models_Via_Machine-Learning_Algorithms
(дата обращения: 25.03.2025).
15. **Merton R. C.** On the Pricing of Corporate Debt: The Risk Structure of Interest Rates // *Journal of Finance*. 1974. Vol. 29, № 2. P. 449–470.
URL: <https://www.jstor.org/stable/2978814>
(дата обращения: 28.02.2025).
16. **Noriega J. P., Rivera L. A., Herrera J. A.** Machine Learning for Credit Risk Prediction: A Systematic Literature Review // *Data*. 2023. Vol. 8, № 11. Art. 169.
URL: <https://www.mdpi.com/2306-5729/8/11/169>
(дата обращения: 10.12.2024).
17. **Pressley J. P.** Why Banks Are Using Advanced Analytics for Faster Fraud Detection // *BizTech Magazine*. 2023. URL: <https://biztechmagazine.com/article/2023/07>
(дата обращения: 10.04.2025).
18. **Varian H. R.** Big Data: New Tricks for Econometrics // *Journal of Economic Perspectives*. 2014. Vol. 28, № 2. P. 3–28.
URL: <https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.28.2.3>
(дата обращения: 10.04.2025).
19. **Vymyatnina Y., Pakhnin M.** Applying Minsky’s Theory to State-Dominated Economies // EUSP Working Paper. 2014. Ec-03/14. URL: https://eusp.org/sites/default/files/archive/ec_dep/wp/Ec-03_14.pdf
(дата обращения: 10.04.2025).
20. **Zhang R., Xue L., Wang Q.** An Ensemble Credit Scoring Model Based on Logistic Regression with Heterogeneous Balancing and Weighting Effects // *SSRN Electronic Journal*. 2022. URL: https://papers.ssrn.com/.../abstract_id=4167821
(дата обращения: 10.12.2024).
21. **Chang V., Sivakulasingam S., Wang H., Wong S. T., Ganatra M. A., Luo J.** Credit Risk Prediction Using Machine Learning and Deep Learning: A Study on Credit Card Customers // *Risks*. 2024. Vol. 12, № 11. Article 174. URL: <https://www.mdpi.com/2227-9091/12/11/174> (дата обращения: 06.02.2025).

Интернет-ресурсы

1. **Federal Deposit Insurance Corporation (FDIC).** Failures and Assistance Transactions – Historical Statistics on Banking, 2001–2025 [Electronic resource]. — URL: <https://banks.data.fdic.gov/explore/failures?>
(дата обращения: 15.02.2025).
2. **Central Bank of the Russian Federation.** *Annual Report 2008.* — Moscow: CBR, 2009.
URL: https://www.cbr.ru/Collection/Collection/File/8322/ar_2008_e.pdf
(дата обращения: 22.12.2024).