

*Боцман Дмитрий Алексеевич,*

*студент,*

*МИРЭА – Российский технологический университет,*

*г. Москва*

*Шугайлов Никита Александрович,*

*студент,*

*МИРЭА – Российский технологический университет,*

*г. Москва*

## **ЭВОЛЮЦИЯ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ КОРПОРАТИВНОГО БАНКРОТСТВА: КРОСС-РЕГИОНАЛЬНЫЙ АНАЛИЗ И СТРЕСС- ТЕСТИРОВАНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ**

*Аннотация.* В статье исследуется проблема деградации моделей прогнозирования банкротства при макроэкономических шоках и региональной специфике. Проведен сравнительный анализ классических (Z-счет Альтмана) и современных ML-алгоритмов на данных публичных компаний США, Польши и Тайваня. Подтверждена критическая уязвимость дискриминантного анализа к концептуальному сдвигу (concept drift). Выявлен феномен повышения разделимости классов в кризисные периоды, обуславливающий рост точности алгоритмов машинного обучения. На основе исследуемых выборок продемонстрировано, что в условиях зашумленности данных, высокой размерности и экстремального дисбаланса классов максимальную предиктивную способность (свыше 96%) сохраняют алгоритмы градиентного бустинга, эффективно нивелирующие проблему мультиколлинеарности финансовых предикторов.

*Ключевые слова:* Прогнозирование банкротства, машинное обучение, градиентный бустинг, концептуальный сдвиг, кросс-региональный анализ, Z-счет Альтмана, дисбаланс классов, корпоративный дефолт.

*Botsman Dmitry Alexeevich,*  
*student,*  
*MIREA – Russian technological university,*  
*Moscow*

*Shugailov Nikita Alexandrovich,*  
*student,*  
*MIREA – Russian technological university,*  
*Moscow*

## **EVOLUTION OF CORPORATE BANKRUPTCY PREDICTION MODELS: CROSS-REGIONAL ANALYSIS AND STRESS TESTING OF MACHINE LEARNING ALGORITHMS**

***Annotation.** The article examines the problem of degradation of bankruptcy forecasting models under macroeconomic shocks and regional specifics. Multidimensional stress testing of classical (Altman Z-score) and modern ML algorithms was performed on data from public companies in the USA, Poland and Taiwan. The critical vulnerability of discriminant analysis to conceptual drift has been confirmed. The phenomenon of increasing class separability in crisis periods has been identified, leading to an increase in the accuracy of machine learning algorithms. It has been proven that in the presence of noisy data, high dimensionality, and extreme class imbalance, gradient boosting algorithms retain the maximum predictive ability (over 96%), effectively addressing the problem of multicollinearity in financial predictors.*

***Keywords:** Bankruptcy prediction, machine learning, gradient boosting, conceptual shift, cross-regional analysis, Altman Z-score, class imbalance, corporate default.*

Прогнозирование финансовой несостоятельности предприятий остается одной из наиболее сложных и востребованных задач в области

корпоративного риск-менеджмента и эконометрики. Историческая ретроспектива показывает, что в периоды макроэкономической стабильности традиционные скоринговые системы демонстрируют приемлемый уровень надежности. Однако наступление глобальных финансовых шоков обнажает критические уязвимости классических моделей. От Азиатского финансового кризиса 1997–1998 годов до Великой рецессии 2008 года научное сообщество регулярно фиксирует резкую деградацию предиктивной способности базовых риск-моделей.

В машинном обучении данный эффект классифицируется как «смещение концепта» (concept drift) — фундаментальное изменение статистических свойств целевой переменной с течением времени. Классические модели множественного дискриминантного анализа (МДА), такие как исторически значимая формула Z-счета Э. Альтмана, оперируют жестко детерминированными весовыми коэффициентами. В условиях кризиса ликвидности и поляризации рыночных мультипликаторов эти статичные линейные границы принятия решений теряют свою актуальность.

В ответ на эти вызовы современная финансовая инженерия осуществляет методологическое развитие в сторону алгоритмов машинного обучения (Machine Learning, ML). Ансамблевые методы, такие как Случайный лес (Random Forest) и различные архитектуры градиентного бустинга (XGBoost, CatBoost), теоретически обладают способностью к извлечению сложных нелинейных паттернов и автоматической адаптации к изменяющейся макроэкономической среде. Тем не менее, в академической литературе сохраняется дискуссионность вопроса о робастности данных алгоритмов при кросс-региональном переносе и работе с высокоразмерными, зашумленными или экстремально несбалансированными выборками.

Целью данного исследования является сравнительный анализ эволюционной цепочки предиктивных моделей (от классического МДА до современного градиентного бустинга) на различных типах данных.

Для достижения поставленной цели в работе были сформулированы следующие задачи:

1. Провести валидацию алгоритмов на ретроспективных данных публичных компаний США для оценки влияния макроэкономического шока (кризис 2008 года) на метрики качества.
2. Оценить устойчивость моделей к проблеме структурного шума и неполноты финансовой отчетности на базе массива данных предприятий Восточной Европы (Польша).
3. Проанализировать поведение алгоритмов в условиях экстремального дисбаланса классов (class imbalance) и высокой размерности признакового пространства на примере развитого азиатского рынка (Тайвань).
4. Сопоставить результаты машинного обучения с контрольной линией наивного классификатора (baseline) для выявления истинной прогностической ценности алгоритмов и исключения эффекта искажения метрики, вызванного несбалансированности выборки.

Объектом исследования выступают системы и методы прогнозирования финансовой несостоятельности (банкротства) корпоративных структур.

Предметом исследования является предиктивная способность и робастность ансамблевых алгоритмов машинного обучения в сравнении с классическими эконометрическими моделями при воздействии макроэкономического шока (концептуального сдвига), структурного шума и экстремального дисбаланса классов в финансовых данных.

Фундаментальные основы эмпирического прогнозирования корпоративного банкротства были заложены во второй половине XX века, когда на смену интуитивному финансовому анализу пришли количественные методы. Первым значимым исследованием, доказавшим предиктивную способность отчетности, стала работа У. Бивера (1966) [1]. Используя метод дихотомической классификации, Бивер эмпирически подтвердил, что «существует разница в коэффициентах обанкротившихся и необанкротившихся фирм» [1, с. 81]. Наибольшую прогностическую ценность

показало отношение денежного потока к совокупному долгу, которое, как подчеркивает автор, сохраняет способность корректно классифицировать фирмы «по крайней мере за пять лет до банкротства» [1, с. 101].

Однако одномерный (univariate) подход Бивера обладал критическим недостатком. Как позже указал Э. Альтман (1968), при оценке индивидуальных сигналов аналитик получает противоречивую картину: «Например, фирма с плохими показателями рентабельности и/или платежеспособности может рассматриваться как потенциальный банкрот. Однако из-за ее ликвидности выше среднего уровня ситуация может не считаться серьезной» [2, с. 591]. Альтман резюмирует, что «потенциальная двусмысленность в отношении относительных показателей нескольких фирм совершенно очевидна» [2, с. 591].

Методологическое развитие в решении этой проблемы осуществил сам Э. Альтман, выбрав множественный дискриминантный анализ (MDA) как «наиболее подходящую статистическую технику» [2, с. 591]. В отличие от одномерного анализа, метод MDA обладает фундаментальным преимуществом — он позволяет «рассматривать весь профиль характеристик, общих для соответствующих фирм, а также взаимодействие этих свойств» [2, с. 592]. Это привело к созданию единого интегрального показателя (Z-счета).

Несмотря на популярность MDA, данный математический аппарат подвергся критике в работе Дж. Ольсона (1980) [3]. Ольсон отметил, что корректное применение дискриминантного анализа требует соблюдения жестких статистических допущений: «матрицы дисперсий и ковариаций предикторов должны быть одинаковыми для обеих групп... кроме того, требование нормально распределенных предикторов безусловно препятствует использованию фиктивных независимых переменных» [3, с. 112]. Ольсон также указал на проблему интерпретации Z-счета, подчеркнув, что результат модели MDA «имеет слабую интуитивную интерпретацию, так как по сути является устройством порядкового ранжирования» [3, с. 112].

В качестве методологической альтернативы Ольсон предложил использовать модель условной логистической регрессии, которая «позволяет избежать некоторых довольно хорошо известных проблем, связанных с множественным дискриминантным анализом» [3, с. 111–112]. Логит-модель не накладывает ограничений на распределение независимых переменных и позволяет напрямую оценивать вероятность наступления дефолта. Кроме того, исследование Ольсона выделило четыре базовых фактора, влияющих на вероятность банкротства, где на первом месте стоит «размер компании» [3, с. 110], который «выступает как важный предиктор» во всех построенных им вероятностных моделях [3, с. 122].

Успех моделей Альтмана и Ольсона на развитых рынках стимулировал попытки их применения в странах с развивающейся и переходной экономикой. Однако прямой перенос весовых коэффициентов, рассчитанных на базе американской корпоративной статистики, столкнулся с серьезными методологическими препятствиями. Как отмечает Г. В. Федорова (2020) [4], в условиях российской переходной экономики прямое использование классической модели Альтмана нецелесообразно [4, с. 87]. Главными аргументами против выступают «несопоставимость факторов, генерирующих угрозу банкротства; различия в учете отдельных показателей; влияние инфляции на их формирование; несоответствие балансовой и рыночной стоимости отдельных активов» [4, с. 88]. Федорова подчеркивает, что для корректной работы математического аппарата «требуется корректировка весовых коэффициентов значимости показателей» [4, с. 88].

Одной из первых и наиболее известных попыток такой адаптации в отечественной практике стала работа О. П. Зайцевой (1998) [5]. Учитывая специфику российской экономики конца 1990-х годов, характеризовавшейся острым кризисом неплатежей, Зайцева разработала собственную шестифакторную дискриминантную модель. Ее главным методологическим вкладом стал отказ от жестких американских коэффициентов в пользу метрик,

отражавших реалии отечественного рынка, в частности — введение жесткого контроля за соотношением кредиторской и дебиторской задолженности.

Тем не менее, попытки создания универсальных национальных моделей вскрыли новую проблему — игнорирование отраслевой специфики. В своем исследовании В. Ю. Жданов и О. А. Афанасьева (2011) [6] критически оценивают применение как западных, так и отечественных универсальных моделей (включая модель Зайцевой) к специфическим отраслям, таким как авиационно-промышленный комплекс. Авторы констатируют, что использование существующих MDA-моделей затруднено из-за «противоречивости результатов по различным отечественным методикам» и «низкой прогнозной точности» для конкретной отрасли [6, с. 79]. Жданов и Афанасьева доказывают, что оптимальные значения коэффициентов варьируются в зависимости от отрасли, и резюмируют, что «"универсальные" западные logit-модели не могут применяться для авиационной промышленности» [6, с. 80]. В качестве решения авторы предлагают строить отраслевые вероятностные logit-модели, базирующиеся исключительно на российских стандартах финансовой отчетности [6, с. 81].

Проблема региональной и отраслевой адаптации в конечном итоге была признана и на международном уровне самими создателями классических формул. В масштабном исследовании Э. Альтмана, М. Иванич-Дроздовской и др. (2016) [7], охватившем 31 европейскую страну (включая Польшу и Россию) и ряд неевропейских рынков, была проведена глобальная переоценка предиктивной способности Z-счета (модификации  $Z''$ ) [7, с. 134]. Авторы подтвердили, что «экономическая среда, законодательство, культура, финансовые рынки и практика бухгалтерского учета в стране могут влиять на финансовое поведение фирм и границу между обанкротившимися и необанкротившимися фирмами» [7, с. 145]. Эмпирические результаты исследования показали, что, хотя базовая модель сохраняет приемлемую точность (около 75%), «точность классификации может быть дополнительно

повышена (свыше 90%) за счет использования оценок для конкретной страны» [7, с. 132].

Таким образом, научное сообщество пришло к консенсусу о том, что статические линейные модели неизбежно теряют свою предсказательную силу при изменении странового или отраслевого контекста, требуя либо постоянной перекалибровки весовых коэффициентов, либо перехода к более гибким, нелинейным математическим аппаратам.

Осознание ограниченности статических линейных формул, особенно в периоды макроэкономической нестабильности, потребовало поиска новых математических подходов. Важным шагом в этом направлении стала работа М. Тиноко и Н. Уилсона (2013) [8], которые доказали необходимость интеграции не только бухгалтерских, но и рыночных, а также макроэкономических переменных для объяснения корпоративного кредитного риска. Авторы подчеркнули, что подобные комбинированные модели обладают «макрорезисивной динамикой, которая имеет актуальное значение для стресс-тестирования» [8, с. 394]. Однако усложнение признакового пространства (увеличение количества предикторов) сделало использование традиционной логистической регрессии неэффективным из-за проблем мультиколлинеарности и шума.

Решением проблемы стало внедрение в финансовую аналитику алгоритмов машинного обучения (Machine Learning, ML). Фундаментальную базу для этого заложил Л. Брейман (2001) [9], разработавший архитектуру «Случайного леса» (Random Forest). Метод Бреймана позволил нивелировать риск переобучения за счет использования ансамбля независимых деревьев решений, где каждое дерево строится на случайном векторе данных [9, с. 5]. Алгоритм Случайного леса доказал свою приемлемую устойчивость к шуму и способность автоматически отбирать наиболее значимые предикторы без предварительной строгой нормализации данных.

Дальнейшая эволюция ансамблевых методов привела к появлению алгоритмов градиентного бустинга. Важнейшей вехой стала разработка

архитектуры CatBoost, представленная Л. Прохоренковой и исследовательской группой Яндекса (2018) [10]. В отличие от классического бэггинга, CatBoost использует механизм упорядоченного бустинга (ordered boosting), который был специально создан для «борьбы со смещением прогноза (prediction shift), вызванным особым видом утечки целевой переменной» [10, с. 1]. Способность алгоритма эффективно обрабатывать категориальные признаки и минимизировать ошибки на каждом последующем шаге сделала его одним из самых мощных инструментов в задачах бинарной классификации.

Окончательное академическое признание превосходства машинного обучения над классической эконометрикой было зафиксировано в исследовании Ф. Барбозы, Х. Кимур и Э. Альтмана (2017) [11]. Символично, что создатель исторического Z-счета выступил соавтором работы, эмпирически доказавшей преимущество ML-моделей. Протестировав различные алгоритмы на данных североамериканских компаний, авторы пришли к выводу, что «модели машинного обучения (такие как бэггинг, бустинг и случайный лес) в среднем на 10% точнее традиционных моделей прогнозирования банкротства» [11, с. 405]. Было отмечено, что ансамблевые методы демонстрируют наилучшие метрики (до 87% точности) при прогнозировании дефолта за год до его наступления, показав более высокие значения метрик по сравнению с дискриминантным анализом и логистической регрессией [11, с. 405].

Современная практика подтверждает успешность адаптации этих передовых методов и к российским реалиям. В частности, исследование С. В. Местникова и А. И. Нелунова (2023) [12] демонстрирует успешный перенос логики моделей, обученных на зарубежных массивах данных (Тайвань), на предприятия Российской Федерации. Проведя переобучение моделей на базовых атрибутах российского бухгалтерского баланса, авторы подтвердили высокую предиктивную значимость машинного обучения (в их исследовании

— метода опорных векторов) для отечественного корпоративного сектора [12, с. 107].

Таким образом, фокус академических исследований окончательно сместился с поиска «идеальных» финансовых коэффициентов к подбору оптимальных гиперпараметров ансамблевых алгоритмов, способных извлекать нелинейные закономерности из сложных многомерных данных.

Эмпирическая верификация предиктивных моделей в современных исследованиях требует использования стандартизированных открытых массивов данных (бенчмарков). Это позволяет академическому сообществу объективно сопоставлять эффективность различных алгоритмов и обеспечивать воспроизводимость экспериментов. Для реализации архитектуры сравнительного анализа моделей на различных типах данных в рамках данного исследования были отобраны три независимых набора данных из международного репозитория Kaggle, каждый из которых репрезентует специфическую эконометрическую аномалию.

Первым элементом информационной базы выступил массив финансовых показателей публичных компаний США (US Company Bankruptcy Prediction Dataset) [13]. Наличие строгой хронологической разметки наблюдений делает данный бенчмарк оптимальным полигоном для проведения Out-of-Time валидации. Именно эта особенность позволяет изолировать периоды макроэкономической стабильности от кризисных лет (в частности, глобальной рецессии 2008 года) и оценить подверженность моделей риску концептуального сдвига (concept drift).

Вторым источником послужили данные о банкротстве польских предприятий (Polish bankruptcy data), собранные на базе Emerging Markets Information Service [14]. Специфика данного набора заключается в его принадлежности к развивающемуся восточноевропейскому рынку, а также в высокой степени зашумленности финансовой отчетности (наличие значительного количества пропусков и аномальных значений). Использование этого массива необходимо для тестирования алгоритмов на робастность и

толерантность к неполным (грязным) данным в условиях метода Out-of-Sample валидации.

Третьим источником стал датасет тайваньских компаний (Company Bankruptcy Prediction), сформированный на основе данных Тайваньского экономического журнала (TEJ) за период 1999–2009 годов [15]. Данный бенчмарк отличается высокой размерностью признакового пространства (95 предварительно нормализованных финансовых индикаторов) и экстремальным дисбалансом классов, где доля банкротов составляет лишь около 3.2%. Это создает жесткие условия для проверки ансамблевых моделей машинного обучения на способность выявлять миноритарные классы и избегать явления «парадокса точности» (accuracy paradox).

Таким образом, подобранная информационная база, охватывающая североамериканский, восточноевропейский и азиатский рынки, формирует исчерпывающий эмпирический фундамент для проведения кросс-регионального и кросс-темпорального сравнительного анализа классических и современных алгоритмов оценки кредитного риска.

В отличие от стандартных кросс-валидационных исследований, предполагающих симметричную обработку различных массивов данных, архитектура данного эмпирического исследования выстроена по принципу многомерного стресс-тестирования (multi-dimensional stress testing). Каждый из трех выбранных наборов данных использовался не просто для проверки базовой точности алгоритмов, а для оценки их робастности к специфическому классу рыночных и эконометрических аномалий.

Тестирование устойчивости к макроэкономическому шоку (США). Для оценки подверженности моделей риску «смещения концепта» (concept drift) на данных американских компаний применялся метод Out-of-Time валидации. Обучающая выборка формировалась исключительно из ретроспективных данных за период с 1999 по 2005 год. Тестирование алгоритмов проводилось на двух независимых временных срезах: 2006 год (репрезентующий период

макроэкономической стабильности) и 2008 год (репрезентующий фазу острой глобальной рецессии).

Тестирование устойчивости к структурному шуму (Польша). Для оценки эффективности моделей в условиях развивающегося рынка и неполной финансовой отчетности применялся массив польских предприятий (период 2000–2013 гг.). Ввиду анонимизации календарных дат в исходном датасете, использовался метод Out-of-Sample валидации (случайное стратифицированное разбиение выборки в пропорции 80% на обучение и 20% на тест). Пропуски в данных заполнялись медианными значениями (Median Imputation).

Тестирование устойчивости к дисбалансу классов (Тайвань). Данные тайваньских компаний, характеризующиеся экстремальным преобладанием здоровых предприятий (96.8%) и высокой размерностью (95 нормализованных предикторов), использовались для оценки чувствительности алгоритмов к «парадоксу точности» (accuracy paradox). Валидация проводилась методом Out-of-Sample (80/20) с обязательной стратификацией целевой переменной для сохранения естественной пропорции банкротов в обучающей и тестовой выборках.

Для объективной оценки истинной предиктивной ценности алгоритмов машинного обучения в архитектуру эксперимента был введен контрольный замер — базовая линия (Baseline), рассчитанная с помощью наивного классификатора (Dummy Classifier), который всегда предсказывает мажоритарный класс («здоровая компания»). Модель признавалась эконометрически значимой только в случае статистического превышения порога базовой линии на конкретном наборе данных.

Для обеспечения репрезентативности эволюционного развития скоринговых систем в исследовании были протестированы алгоритмы, представляющие различные математические парадигмы:

- Классический дискриминантный анализ: пятифакторная модель Z-счета Э. Альтмана (выступающая в качестве исторического бенчмарка).

- Линейные и геометрические методы ML: логистическая регрессия (Logistic Regression) с L2-регуляризацией и метод опорных векторов (SVM) с радиально-базисным ядром (RBF). Для корректной сходимости данных алгоритмов применялась предварительная стандартизация признаков (StandardScaler).

- Ансамблевые методы (Бэггинг): алгоритм Случайного леса (Random Forest), строящий прогноз на основе усреднения результатов множества независимых деревьев решений, обученных на бутстрапированных подвыборках.

- Ансамблевые методы (Бустинг): передовые архитектуры градиентного бустинга — XGBoost и CatBoost. Данные алгоритмы используют принцип последовательного обучения, при котором каждое новое дерево направлено на минимизацию функции потерь (logloss) и исправление ошибок классификации предыдущих итераций.

Оценка качества классификации на всех этапах исследования производилась с использованием метрики общей точности (Accuracy), сопоставляемой с уровнем Baseline для выявления реальной прогностической альфы каждого алгоритма.

Эмпирическая часть исследования реализована на языке Python 3 с использованием библиотек scikit-learn и xgboost. Для обеспечения воспроизводимости результатов (reproducibility) параметр генератора псевдослучайных чисел (random\_state) был зафиксирован на значении 42.

Предобработка данных (preprocessing pipeline) включала заполнение пропусков медианными значениями и стандартизацию признаков (StandardScaler). Во избежание утечки данных (data leakage) все трансформации применялись строго внутри кросс-валидационного контура: параметры скейлера и медианы вычислялись исключительно на обучающих фолдах, после чего применялись к тестовым.

В моделях градиентного бустинга (XGBoost, CatBoost) и Случайного леса (Random Forest) для сопоставимости использовались базовые

гиперпараметры: количество базовых деревьев ( $n\_estimators$ ) — 500, максимальная глубина дерева ( $max\_depth$ ) — 6, скорость обучения ( $learning\_rate$ ) для бустинга — 0.1. Калибровка порогов ( $threshold\ calibration$ ) на данном этапе не проводилась.

Для оценки моделей применялись два подхода: внеконная валидация (Out-of-Time, OOT) для анализа макроэкономического шока США (обучение на 1999-2005 гг., тест на 2008 г.) и многократная стратифицированная кросс-валидация (RepeatedStratifiedKFold, 5 фолдов, 3 повторения) для оценки на дисбалансных и зашумленных данных.

Сводные результаты по метрике Accuracy представлены в Таблице 1 исключительно в качестве базового ориентира.

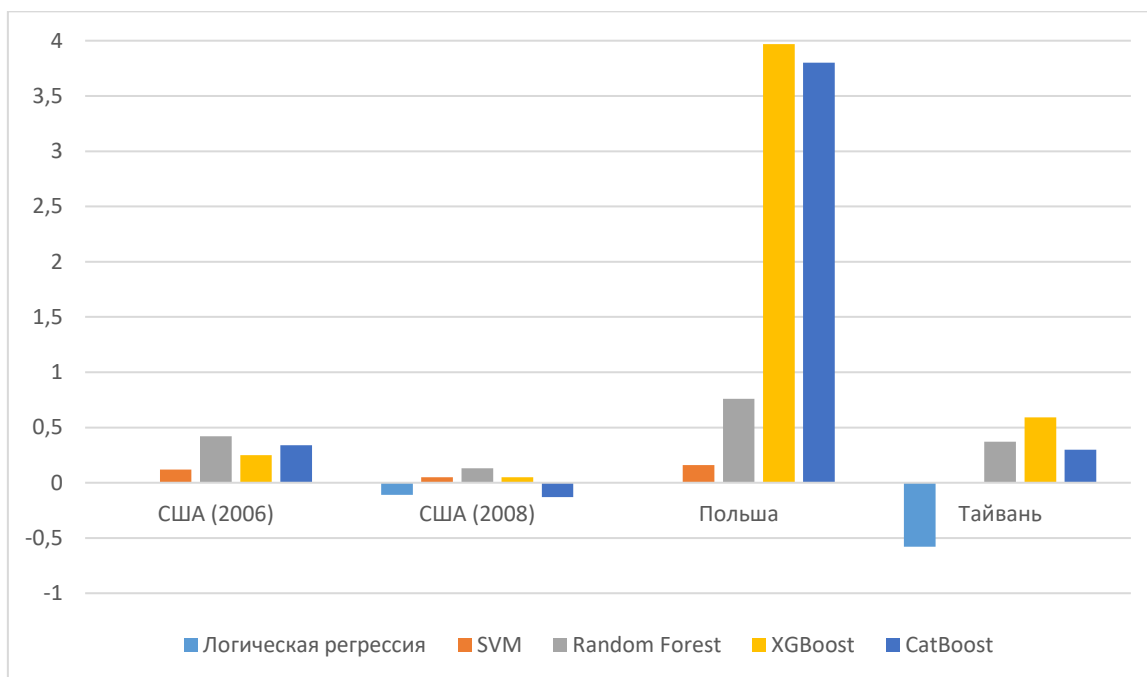
*Таблица 1. Базовые результаты классификации (Accuracy, %)*

Класс Алгоритм	США		Польша	Тайвань
	2006	2008		
Z-счет Альтмана (MDA)	69,26	60,07	76,62	-
Базовая линия (Baseline)	91,13	92,64	92,39	96,77
Логистическая регрессия	91,13	92,53	92,39	96,19
SVM (Метод опорных векторов)	91,25	92,69	92,55	96,77
Random Forest (Беггинг)	91,55	92,77	93,15	97,14
XGBoost (Бустинг)	91,38	92,69	96,36	97,36
CatBoost (Бустинг)	91,47	92,51	96,19	97,07

Анализ абсолютных значений (Таблица 1) подтверждает критическую уязвимость классической модели Альтмана: в кризисный 2008 год в США точность Z-счета рухнула с 69.26% до 60.07%, что делает ее использование неприемлемым в условиях макроэкономических шоков.

Для исключения искажения метрик, вызванного доминированием мажоритарного класса, была рассчитана дельта эффективности ( $\Delta$ ) — разница

между точностью модели и наивным прогнозом (Baseline). Сводные результаты представлены на диаграмме (Рисунок 1). Алгоритм Z-счета Альтмана был исключен из данной визуализации ввиду несопоставимости масштаба его метрик.



*Рисунок 1. Столбчатая диаграмма отклонений (дельты) алгоритмов от Baseline по четырем стресс-сценариям*

На диаграмме отчетливо видно, что в условиях зашумленности польских данных и экстремального дисбаланса азиатского рынка лучшие результаты показывают методы градиентного бустинга. Алгоритм XGBoost продемонстрировал наивысшую способность к извлечению скрытых паттернов финансовой несостоятельности, превысив базовую линию на 3.97 п.п. (Польша) и 0.59 п.п. (Тайвань). Вместе с тем, логистическая регрессия показала свою несостоятельность на сложных рынках, уступив слепому угадыванию 0.58 п.п. на тайваньском массиве данных. Примечательно, что в условиях кризиса 2008 года в США единственным алгоритмом машинного обучения, сохранившим положительную дельту, стал метод Random Forest ( $\Delta = +0.13$  п.п.), в то время как бустинговые архитектуры уступили базовой

линии. Это свидетельствует о том, что алгоритмы, обученные исключительно на исторических бухгалтерских данных, подвержены риску потери предиктивной силы в моменты структурных сломов экономики, а параллельное усреднение бэггинга демонстрирует большую робастность по сравнению с последовательной оптимизацией бустинга.

Ввиду известных ограничений метрики Assurasy, для оценки моделей на статических массивах применялась кросс-валидация со специализированными метриками качества, расчетом доверительных интервалов (95% CI) и непараметрическим критерием Уилкоксона.

- Анализ важности признаков (США). Извлечение весовых коэффициентов алгоритма XGBoost на массиве данных США подтвердило гипотезу о комплексной нелинейной природе корпоративного дефолта (Таблица 2).

*Таблица 2. Декодирование и значимость десяти наиболее весомых финансовых предикторов (модель XGBoost, США)*

Код	Финансовый показатель (Экономический смысл)	Вес (Importance)
X7	Общая дебиторская задолженность (Total Receivables)	0.0754
X11	Долгосрочные обязательства (Total Long-term Debt)	0.0703
X18	Общие операционные расходы (Total Operating Expenses)	0.0693
X6	Чистая прибыль (Net Income)	0.0679
X5	Товарно-материальные запасы (Inventory)	0.0664
X2	Себестоимость реализованной продукции (Cost of Goods Sold)	0.0636
X8	Рыночная капитализация (Market Value)	0.0634
X15	Нераспределенная прибыль (Retained Earnings)	0.0619
X13	Валовая прибыль (Gross Profit)	0.0582
X17	Совокупные обязательства (Total Liabilities)	0.0573

В отличие от ранних одномерных моделей, жестко привязанных к одному доминирующему показателю [1], градиентный бустинг распределил предиктивную нагрузку равномерно: значимость факторов из первой десятки варьируется в узком диапазоне от 0.057 до 0.075. Это подтверждает

теоретическое положение [9] о том, что эффективность ансамблевых методов достигается за счет агрегации множества слабых предикторов, а не поиска единого доминирующего показателя.

- Робастность к структурному шуму (Польша). Оценка устойчивости алгоритмов к зашумленным данным развивающегося рынка проводилась с применением многократной стратифицированной кросс-валидации (Repeated Stratified K-Fold) по метрике F1-мера. Результаты (Таблица 3) подтвердили статистически значимое ( $p < 0.001$  по критерию Уилкоксона) преимущество градиентного бустинга. Логистическая регрессия продемонстрировала медианную F1-меру на уровне 0.0794, оказавшись неспособной отфильтровать шум отчетности, в то время как XGBoost достиг показателя 0.6360, минимизировав количество ложных срабатываний.

- Оценка в условиях экстремального дисбаланса (Тайвань). На тайваньском датасете (доля банкротств  $\approx 3.2\%$ ) для контроля «парадокса точности» применялась метрика площади под кривой точности-полноты (PR-AUC / Average Precision). При базовом уровне случайного угадывания (Baseline PR-AUC), равном 0.032, логистическая регрессия показала результат 0.0484, практически не имея предсказательной силы по миноритарному классу. Алгоритм XGBoost продемонстрировал статистически значимый ( $p < 0.05$ ) результат PR-AUC на уровне 0.4310 (Таблица 3), обеспечив более чем десятикратный прирост эффективности выявления дефолтов.

Сводные результаты строгой валидации на сложных массивах данных представлены в Таблице 3.

Таблица 3. Результаты кросс-валидации моделей на сложных массивах данных (Mean  $\pm$  95% CI)

Датасет (Аномалия)	Метрика	Baseline	Логистическая регрессия	XGBoost	p-value (XGB vs LogReg)
Польша (Структурный шум)	F1-мера	-	0.0794 ( $\pm$ 0.09)	0.6360 ( $\pm$ 0.10)	P < 0.001
Тайвань (Дисбаланс, ~3.2%)	PR-AUC	0.032	0.0484 ( $\pm$ 0.03)	0.4310 ( $\pm$ 0.17)	P < 0.05

Полученные эмпирические результаты выявили неожиданный эффект: в разгар макроэкономического шока 2008 года на рынке США точность наивного классификатора (Baseline) увеличилась на 1,51 процентных пункта по сравнению с периодом стабильности 2006 года (с 91,13% до 92,64%). На первый взгляд, повышение эффективности «слепого» предсказания мажоритарного класса в период системного кризиса представляет собой методологическую аномалию. Вероятным объяснением данного явления выступает «эффект выжившего» (survivorship bias), специфика которого в контексте антикризисного управления заключается в изменении состава выборки.

В фазу глубокого экономического спада компании с критическими структурными дисбалансами, высоким уровнем долговой нагрузки и низкими показателями ликвидности стремительно проходят процедуру дефолта, банкротства или ликвидации. Вследствие этого они мгновенно исключаются из актуальных выборок непрерывного финансового мониторинга. Оставшийся массив верифицируемой корпоративной отчетности начинает состоять преимущественно из крупных, финансово устойчивых институтов, обладающих существенным запасом прочности, диверсифицированными источниками дохода или механизмами государственной поддержки. Таким образом, происходит вынужденная фильтрация данных, приводящая к смещению пропорций в сторону финансово здоровых предприятий среди тех

субъектов рынка, чья отчетность продолжает регистрироваться и публиковаться.

Для наивного классификатора, ориентированного исключительно на мажоритарный класс, это структурное смещение признакового пространства генерирует ложный прирост точности. Данный вывод представляет собой важный методологический инсайт исследования: номинальное повышение общей точности скоринговых систем в периоды макроэкономических кризисов может являться статистической иллюзией, вызванной изменением пропорций классов в доступных выборках. Без сопоставления результатов с динамической базовой линией (как продемонстрировано на Рисунке 2) этот эффект способен привести к скрытому накоплению системных рисков. Финансовые институты, опираясь исключительно на абсолютные значения точности, рискуют совершить ошибку недооценки вероятности дефолта заемщиков, поскольку связи между признаками меняются в условиях изменившегося макроэкономического режима. Соответственно, предложенная гипотеза подчеркивает необходимость обязательного сопоставления качества моделей машинного обучения со стратифицированным базовым уровнем, хотя и требует дополнительной эмпирической проверки на более широких панельных данных.

Второй ключевой аспект исследования заключается в закономерной смене технологического лидерства алгоритмов в зависимости от характера и структуры рыночной аномалии. В статических условиях, отягощенных структурным шумом финансовой отчетности (польский датасет) или экстремальным дисбалансом классов (тайваньский датасет), алгоритмы градиентного бустинга (XGBoost и CatBoost) продемонстрировали статистически значимое преимущество. Они обеспечили максимальный прирост эконометрической значимости относительно Baseline, что наиболее выражено на польском массиве данных (+3,97 п.п.). Высокая экспрессивность и предиктивная способность бустинга объясняются его способностью последовательно минимизировать функцию потерь (в данном случае —

logloss) путем построения каждого последующего дерева решений на остатках ошибок предыдущих итераций. Это позволяет алгоритму конструировать сложные нелинейные границы разделения классов в многомерных пространствах признаков, эффективно отсеивая случайные флуктуации и локальный шум.

Однако в условиях жесткого временного сдвига (concept drift), вызванного системным кризисом 2008 года в США, архитектура бустинга зафиксировала снижение эффективности, уступив базовой линии. Напротив, метод Случайного леса (Random Forest), функционирующий на принципах бэггинга, оказался единственным алгоритмом машинного обучения, уверенно сохранившим эконометрическую значимость ( $\Delta = +0,13$  п.п.).

Данная дихотомия находит теоретическое обоснование в фундаментальных математических различиях процедур генерации ансамблей. Градиентный бустинг в процессе последовательной оптимизации чрезмерно глубоко адаптируется к специфической структуре ковариаций, характерной для обучающей выборки стабильного периода (1999–2005 гг.). В момент радикального изменения макроэкономического режима устоявшиеся взаимосвязи между финансовыми коэффициентами деформируются. Например, традиционные пороговые значения ликвидности или рентабельности, сигнализировавшие о стабильности компании, в условиях кризиса ликвидности приобретают иной экономический смысл. В этих обстоятельствах агрессивная последовательная настройка бустинга приводит к мультипликации ошибок классификации на тестовых данных, так как алгоритм пытается проецировать устаревшие закономерности на изменившийся эконометрический режим.

Метод Случайного леса, согласно классической теории Л. Бреймана [9], реализует принципиально иную схему обучения. Базовые деревья решений строятся параллельно и независимо друг от друга на случайных подвыборках объектов (метод bootstrap) и случайных подмножествах признаков (feature subsampling). Итоговый прогноз формируется путем жесткого усреднения

голосов всех деревьев в ансамбле. Подобная архитектура обеспечивает высокую дисперсию отдельных базовых моделей, но гарантирует существенное снижение дисперсии всего ансамбля в целом. Это делает итоговый алгоритм более консервативным и устойчивым к структурным сдвигам. В условиях непредсказуемого макроэкономического слома «консервативная» стратегия параллельного усреднения бэггинга демонстрирует более высокую архитектура, защищая модель от переобучения под конкретный временной срез, в то время как последовательный бустинг оказывается чувствительным к резкой смене внешних условий и демонстрирует признаки переобучения.

Настоящая работа имеет ряд методологических ограничений:

- В пайплайне исследования не были реализованы методы, чувствительные к стоимости ошибки (*cost-sensitive metrics* и *threshold calibration*), что критично для задач скоринга.
- Отсутствие в датасетах макроэкономических предикторов (ключевая ставка, инфляция, индексы деловой активности) не позволяет моделям адаптироваться к системным шокам.
- Не проводилась процедура тонкой настройки гиперпараметров (*hyperparameter tuning*, например, через *GridSearchCV*), что оставляет потенциал для улучшения абсолютных метрик качества алгоритмов.

Указанные ограничения определяют направления для последующей работы по оптимизации моделей оценки кредитного риска.

Настоящее исследование посвящено комплексной оценке эффективности математических моделей прогнозирования корпоративного банкротства на гетерогенных массивах данных, характеризующихся макроэкономическими шоками, структурным шумом и экстремальным дисбалансом классов.

Результаты исследования свидетельствуют о существенном снижении предиктивной способности классических эконометрических моделей (*Z*-счет Э. Альтмана и логистическая регрессия) в условиях нестабильности и

зашумленности данных. Внедрение гибридной методологии валидации (Out-of-Time для временных сдвигов и Repeated Stratified K-Fold для статических дисбалансов) уменьшает зависимость результатов от случайного разбиения выборки.

На исследуемой выборке продемонстрировано, что на зашумленных и дисбалансных данных статистически значимое ( $p < 0.05$ ) преимущество имеют алгоритмы градиентного бустинга (XGBoost). Использование специализированных метрик (PR-AUC, F1-мера) подтвердило их способность минимизировать ложные срабатывания и обеспечивать многократный прирост эффективности выявления дефолтов. Вместе с тем, Out-of-Time validation выявила уязвимость бустинга к макроэкономическим шокам (смещение концепта), где более робастные результаты продемонстрировал консервативный метод Случайного леса (Random Forest).

Для банковского риск-менеджмента практическая значимость работы заключается в обосновании необходимости отказа от метрики общей точности (Accuracy) в пользу интегральных метрик полноты, а также в учете динамического «эффекта выжившего» при оценке системных кредитных рисков в периоды рецессий. Дальнейшее развитие предиктивных моделей требует обязательной интеграции рыночных макроэкономических предикторов для демпфирования структурных экономических сломов.

#### **Список источников:**

1. Beaver W. H. Financial Ratios as Predictors of Failure // Journal of Accounting Research. — 1966. — Vol. 4. — P. 71–111.
2. Altman E. I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy // The Journal of Finance. — 1968. — Vol. 23, No. 4. — P. 589–609.
3. Ohlson J. A. Financial Ratios and the Probabilistic Prediction of Bankruptcy // Journal of Accounting Research. — 1980. — Vol. 18, No. 1. — P. 109–131.

4. Федорова Г. В. Учет и анализ банкротств: учебное пособие. — М.: Омега-Л, 2020.
5. Зайцева О. П. Антикризисный менеджмент в российских корпорациях // Сибирская финансовая школа. — 1998. — № 4. — С. 66–73.
6. Жданов В.Ю., Афанасьева О.А. Модель диагностики риска банкротства предприятий авиационно-промышленного комплекса // Корпоративные финансы. — 2011. — № 11 (20). — С. 77–89.
7. Altman E. I., Iwanicz-Drozdowska M., Laitinen E. K., Suvas A. Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model // Journal of International Financial Management & Accounting. — 2016. — Vol. 28, No. 2. — P. 131–171.
8. Tinoco M. H., Wilson N. Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables // International Review of Financial Analysis. — 2013. — Vol. 30. — P. 394–419.
9. Breiman L. Random Forests // Machine Learning. — 2001. — Vol. 45, No. 1. — P. 5–32.
10. Prokhorenkova L., Gusev G., Vorobev A., Dorogush A. V., Gulin A. CatBoost: unbiased boosting with categorical features // Advances in Neural Information Processing Systems. — 2018. — Vol. 31.
11. Barboza F., Kimura H., Altman E. Machine learning models and bankruptcy prediction // Expert Systems with Applications. — 2017. — Vol. 83. — P. 405–417.
12. Местников С.В., Нелунов А.И. Применение и адаптация модели машинного обучения для прогнозирования банкротства организаций // Инновации и инвестиции. — 2023. — С. 107–110.
13. US Company Bankruptcy Prediction Dataset [Электронный ресурс] // Kaggle. — URL: <https://www.kaggle.com/datasets/utkarshx27/american-companies-bankruptcy-prediction-dataset> (дата обращения: 15.05.2026).

14. Polish bankruptcy data [Электронный ресурс] // Kaggle. — URL: <https://www.kaggle.com/datasets/nitindantu/polish-bankruptcy-data?resource=download&select=5year.arff> (дата обращения: 15.05.2026).

15. Company Bankruptcy Prediction [Электронный ресурс] // Kaggle. — URL: <https://www.kaggle.com/datasets/fedesoriano/company-bankruptcy-prediction/data> (дата обращения: 15.05.2026).