Ю. Ю. Феста
независимый исследователь (Москва)

ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ МЕТРИК ТОЧНОСТИ БИНАРНОЙ КЛАССИФИКАЦИИ ПРИ УВЕЛИЧЕНИИ МИНОРИТАРНОГО КЛАССА В СЛУЧАЕ ДИСБАЛАНСА

Введение

Оценка вероятности дефолта крупных заемщиков, несмотря на долгую историю исследования данной тематики, остается актуальной задачей для банков и государственных регуляторов. В 2023 г. Центральный банк Российской Федерации опубликовал консультационный доклад, посвященный использованию методов машинного обучения и искусственного интеллекта в финансовой сфере (Банк России, 2023). В нем рассматриваются перспективные возможности применения технологий ML, такие как повышение точности оценки рисков и улучшение обработки финансовых данных. В документе также отмечены потенциальные проблемы, включая зависимость от качества исходных данных и необходимость высокого уровня компетенций у специалистов, работающих с моделями. Вопрос регулирования применения машинного обучения в финансовом секторе активно обсуждается и на международном уровне (Manera, 2024). Современные методы машинного обучения и искусственного интеллекта позволяют анализировать огромные массивы данных, однако на практике аналитики и исследователи нередко сталкиваются с проблемой недостаточного объема данных: отсутствием репрезентативных выборок, необходимых для обучения и валидации сложных моделей. Для качественного обучения сложных моделей, таких как градиентный бустинг или нейросети, требуются тысячи и десятки тысяч наблюдений, предварительно отобранных аналитиками, к тому же в случае анализа низкодефолтных портфелей (LDP) проблема формирования выборки становится еще более сложной. Сразу отпадает такой метод, как андерсемплинг, при котором мы убираем наблюдения мажоритарного класса для балансировки выборки. Использование алгоритмов оверсемплинга, например SMOTE (Chawla et al., 2003), на критически малых данных в случае низкодефолтных портфелей, может усиливать искажение распределений из-за дублирования наблюдений. SMOTE создает синтетические наблюдения путем линейной интерполяции между представителями миноритарного класса. Однако при крайне малом размере миноритарного класса выбор соседей становится статистически ненадежным. Если дефолты представлены единичными наблюдениями, синтетические примеры, генерируемые на их основе, могут привести к генерации точек в областях, где реальные данные отсутствуют, или усилить корреляции, не отражающие истинные экономические зависимости. В сообществе исследователей нет консенсуса относительно применения подобных алгоритмов. В одних случаях их использование может привести к улучшению качества работы м одели (Xing et al., 2024, Pradeep et al., 2024), а в других случаях — наоборот, ухудшить качество работы моделей. В работе (Goorbergh et al., 2022) показано, что SMOTE критически ухудшил метрику AUC-ROC для алгоритма логистической регрессии в задаче бинарной классификации.

DOI 10.31085/1814-4802-2025-21-3-168-51-60
© Ю. Ю. Феста, 2025

Генерация исходного сбалансированного набора данных

Методология исследования включает генерацию исходного набора данных с равным балансом классов, после чего требуемый дисбаланс достигается за счет контролируемого удаления объектов целевого класса (дефолтов) с учетом типа пропусков в данных. На втором этапе производится сравнение метрик классификации (*Accuracy, Precision, Recall, F1*) и их доверительных интервалов при различных уровнях дисбаланса, что позволяет оценить влияние сокращения доли дефолтов на качество модели.

Для создания меток целевого класса изначально рассчитывается параметр *Score*. Формально его оценка определяется как:

$$Score_j = \sum_{i=1}^{n} X_{ij} + w \cdot \omega_i, \tag{1}$$

где $X_{ij} \sim N(0,1)$ представляют собой базовые факторы кредитного риска — независимые стандартные нормальные случайные величины. Шумовая компонента $w_i \sim U(0,1)$ представляет собой равномерно распределенную случайную величину, вклад которой регулируется весовым коэффициентом $w_i = (0{,}1; 0{,}5; 1{,}0)$.

Для присвоения метки целевого класса используется пороговое значение величины *Score*, равное медиане этой величины по всем наблюдениям в выборке:

$$Y_j = 0 \mid Score_j \leqslant Median(Score_j), \text{ else } Y_j = 1. \tag{2}$$

Предположение о стандартном нормальном распределении факторов X_{ij} соответствует практикам машинного обучения, так как стандартизация входных данных (z-нормализация) устраняет масштабные различия между признаками, повышая стабильность алгоритмов, а также обеспечивает интерпретируемость факторов в единицах стандартного отклонения.

Генерация пропусков в данных

В научной литературе принято разделять полностью случайные пропуски (MCAR), случайные пропуски, связанные с наблюдаемыми данными (MAR), и неслучайные пропуски, зависящие от ненаблюдаемых факторов (MNAR). Изначально эта классификация пропусков относилась к объясняющим переменным (Gomer, Yuan, 2023), но в данной работе предлагается экстраполировать эту классификацию на все наблюдения. При создании синтетических данных особую сложность составляет задача генерирования случайных и абсолютно случайных пропусков. Любой алгоритм будет детерминирован, и каждый метод можно считать условно случайным. При этом мы имеем две категории случайности, которые стоит различать по степени детерминированности: не детерминировано, слабо детерминировано. Метрология исследования предполагает, что наблюдение считается пропущенным абсолютно случайно (MCAR), если его отсутствие не зависит ни от наблюдаемых, ни от скрытых переменных в данных, т. е. степень детерминированности низкая. Иными словами, функция, генерирующая пропуск, не должна принимать как аргумент значения показателей или их производные значения. Наблюдение считается пропущенным случайно (MAR), если вероятность его отсутствия косвенно связана с наблюдаемыми переменными, но не выражается явной математической зависимостью — такие пропуски могут объясняться отдельными признаками набора данных, но природа пропусков носит статистический, а не математически детерминированный характер.

— Генерация абсолютно случайных пропусков по заемщикам реализована через последовательные испытания Бернулли с вероятностью удаления наблюдения $p = 0{,}005$. На каждом шаге случайным образом исключались данные, что соответствует сценарию технических сбоев или случайной потери информации (например, пропуск в загрузке кредитной истории). Процесс повторялся до достижения заданной доли дефолтов в выборке, сохраняя случайный характер пропусков без зависимости от других переменных.
— Для генерации случайных пропусков алгоритм удалял каждое i-е наблюдение в цикле, где i динамически определялось текущим размером данных. Это моделирует ситуации, когда отсутствие информации связано с наблюдаемыми параметрами заемщика. Процедура продолжалась до достижения целевой доли дефолтов, создавая систематические пропуски, зависящие от наблюдаемого признака, но не зависящие от параметра *Score*.

Оба метода позволяют изучать влияние пропусков на предсказательную силу моделей вероятности дефолта. MCAR имитирует абсолютно недетерминированные случаи пропусков, тогда как MAR отражает более интерпретируемые сценарии, где пробелы в данных связаны с доступными характеристиками заемщиков.

Метрики качества алгоритмов классификации

Задача моделирования вероятности дефолта относится к задачам бинарной классификации, где ключевым инструментом оценки качества алгоритма выступает матрица ошибок (Raschka, Mirjalili, 2023), представленная в табл. 1. Эта матрица сопоставляет истинные и предсказанные классы, разделяя результаты на четыре категории:

True Positive (TP) — корректно предсказанные дефолты,

True Negative (TN) — корректно предсказанные недефолты,

False Positive (FP) — ложные срабатывания (недефолты, ошибочно классифицированные как дефолты),

False Negative (FN) — пропущенные дефолты.

Таблица 1

Матрица ошибок классификации

	Дефолт (предсказано)	Не дефолт (предсказано)
Дефолт (факт)	TP	FP
Не дефолт (факт)	FN	TN

На основе матрицы рассчитываются метрики, отражающие разные аспекты эффективности модели: как *Accuracy, Precision, Recall* и *F*1-мера. В данном исследовании рассматриваются все четыре метрики, что обусловлено необходимостью баланса между интерпретируемостью, экономической значимостью и адекватностью в условиях дисбаланса классов, характерного для кредитных данных.

Accuracy характеризует долю верных предсказаний относительно всех объектов выборки (3). Эта метрика может быть некорректной при значительном дисбалансе классов. Несмотря на простоту интерпретации, данная метрика может вводить в заблуждение при сильном дисбалансе классов (например, 5% дефолтов). Если модель классифицирует все наблюдения как недефолты, значение метрики составит 95%, что является неадекватной оценкой качества:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{3}$$

Precision — это доля правильных предсказаний класса от общего числа случаев истинных значений целевого класса. Высокий показатель данной метрики указывает на минимизацию доли ложных срабатываний (FP), что может быть критично для снижения операционных издержек (например, необоснованный отказ в кредите надежным заемщикам), и соответствия регуляторным требованиям (недопущение неоправданных резервов под потери):

$$Precision = \frac{TP}{TP + FP}. \tag{4}$$

Recall — это способность модели выявлять истинные дефолты. Максимизация данной метрики снижает риск пропуска дефолтов (FN), что напрямую влияет на финансовую устойчивость банка (избежание убытков):

$$Recall = \frac{TP}{TP + FN}. \tag{5}$$

F1-мера — гармоническая средняя *Precision* и *Recal*. Данная метрика наиболее часто встречается в публикациях по оценке алгоритмов бинарной классификации:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}. \tag{6}$$

Доверительные интервалы метрик качества алгоритмов классификации

Использование доверительных интервалов в оценке качества алгоритмов бинарной классификации критически важно, особенно в контексте низкодефолтных портфелей, где дисбаланс классов и малое число положительных примеров создают высокую неопределенность. Доверительные интервалы позволяют оценить диапазон, в котором с заданной вероятностью (например, 95%) находится истинное значение метрик, таких как *Precision*, *Recall* или $F1$-мера, что исключает переоценку эффективности модели. В случае низкодефолтных портфелей, где доля дефолтов может составлять менее 1%, даже незначительные изменения в данных (например, пропуск одного дефолта) приводят к значительным колебаниям метрик. В данной работе для каждой из метрик алгоритма классификации рассчитывались следующие доверительные интервалы: нормальное приближение, интервалы Агрести–Коулли (Agresti–Coull), Уилсона (Wilson), Бета (Clopper–Pearson) и Джеффреса (Jeffreys).

Доверительные интервалы на основе нормального распределения могут быть эффективны при больших объемах данных, когда в соответствии с центральной предельной теоремой обеспечивается приближение к нормальности выборочных оценок. Однако доверительные интервалы такого рода могут давать неточные результаты для малых выборок или даже отрицательные границы интервалов при сильной асимметрии. Формула для расчета нижней и верхней границы для нормальной аппроксимации:

$$CI_l^N = \left(S / n\right) - \gamma_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\left(\frac{S}{n}\right) \cdot \left[1 - \left(\frac{S}{n}\right)\right]}{n}}, \tag{7}$$

$$CI_u^N = \left(S / n\right) + \gamma_{\alpha/2} \cdot \sqrt{\frac{\left(\frac{S}{n}\right) \cdot \left[1 - \left(\frac{S}{n}\right)\right]}{n}}, \tag{8}$$

где n — число испытаний; S — число успехов; $\gamma_{\alpha/2}$ — квантиль нормального распределения с уровнем значимости $\alpha/2$.

Особенность доверительных интервалов Уилсона заключается в смещении оценок за счет учета асимметрии распределения при сохранении адекватной ширины интервала даже в условиях, где нормальное приближение дает нереалистичные результаты (Lott, Reiter, 2022). Формулы для расчета нижней и верхней границы интервала Уилсона:

$$CI_l^W = \frac{S + \left(\gamma^2_{\alpha/2} / 2\right)}{n + \gamma^2_{\alpha/2}} - \gamma_{\alpha/2} \cdot \sqrt{\frac{SF / n^2 + \gamma^2_{\alpha/2} / 4}{n + \gamma^2_{\alpha/2}}}, \tag{9}$$

$$CI_l^W = \frac{S + \left(\gamma^2_{\alpha/2} / 2\right)}{n + \gamma^2_{\alpha/2}} + \gamma_{\alpha/2} \cdot \sqrt{\frac{SF / n^2 + \gamma^2_{\alpha/2} / 4}{n + \gamma^2_{\alpha/2}}}, \tag{10}$$

где $\gamma_{\alpha/2} = 2$; F — число неудач.

Доверительные интервалы Клоппера—Пирсона (Бета-интервалы) — это точный метод для биномиальных пропорций, гарантирующий, что вероятность покрытия истинного значения не ниже заявленного уровня значимости (Puza, O'Neill, 2006). Формулы для расчета нижней и верхней границы Бета-интервала:

$$CI_l^{CP} = \frac{1}{1 + \frac{n - S + 1}{S} \cdot F_{2(n-s+1), 2S, \alpha/2}}, \tag{11}$$

$$CI_u^{CP} = \frac{\frac{s+1}{n-S} \cdot F_{2(s+1), 2S, \alpha/2}}{1 + \frac{s+1}{n-S} \cdot F_{2(s+1), 2S, \alpha/2}}, \tag{12}$$

где $F_{u,v,\gamma}$ — это значения F-распределения со степенями свободы u и v с уровнем значимости γ.

Доверительные интервалы Агрести—Коули дополняют стандартное нормальное приближение для случая ограничения на отрицательные значения внутри интервала (Kim, Jang, 2022). Формулы для расчета нижней и верхней границы интервала Агрести—Коули:

$$CI_l^{AC} = \frac{S + \left(\gamma^2_{\alpha/2} / 2\right)}{n + \gamma^2_{\alpha/2}} - \gamma_{\alpha/2} \cdot \frac{\sqrt{(SF) \cdot \left[1 + \left(\gamma^2_{\alpha/2} / 2\right)\right] + \left(\gamma^4_{\alpha/2} / 4\right)}}{n + \gamma^2_{\alpha/2}}, \tag{13}$$

$$CI_l^{AC} = \frac{S + \left(\gamma^2_{\alpha/2} / 2\right)}{n + \gamma^2_{\alpha/2}} + \gamma_{\alpha/2} \cdot \frac{\sqrt{(SF) \cdot \left[1 + \left(\gamma^2_{\alpha/2} / 2\right)\right] + \left(\gamma^4_{\alpha/2} / 4\right)}}{n + \gamma^2_{\alpha/2}}. \tag{14}$$

Доверительные интервалы Джеффриса основаны на байесовском подходе с использованием бета-распределения, что делает расчет доверительных интервалов устойчивым для малых выборок (Brown et al., 2001). Формула для расчета нижней и верхней границы интервала Джеффреса:

$$CI_l^J = Beta\left(\frac{\alpha}{2};\ S+0{,}5;\ n-S+0{,}5\right), \tag{15}$$

$$CI_u^J = Beta\left(1-\frac{\alpha}{2};\ S+0{,}5;\ n-S+0{,}5\right), \tag{16}$$

где $Beta(\alpha, \alpha_1, \alpha_2)$ — квантиль Бета-распределения с параметрами α_1 и α_2.

Для каждого из типов распределений значения S (число успехов) и F (число неудач) для каждой из метрик следует из матрицы ошибок классификации и оцениваемых метрик. Эти значение показаны в табл. 2.

Таблица 2

Парметры S и F для каждой из метрик классификации

Метрика	Число успехов	Число неудач
Accuracy	$TP + TN$	$FP + FN$
Precision	TP	FP
Recall	TP	FN
$F1$	$2*TP$	$FN+FP$

Исследуемые наборы данных и их спецификации

В данном исследовании оценивались четыре метрики качества алгоритма классификации — *Accuracy*, *Precision*, *Recall* и $F1$ и пять типов их доверительных интервалов — нормальное, Агрести–Коулли (Agresti–Coull), Уилсона (Wilson), Бета (Clopper–Pearson) и Джеффреса (Jeffreys). При генерации данных учитывались три весовых коэффициента для шума (0, 0,5, 1,0). Были рассмотрены выборки двух размеров: 100 наблюдений и 20 000 наблюдений. Для получение необходимой доли дефолта учитывалось два типа пропусков в данных — MCAR и MAR. Также учитывались четыре спецификации для объясняющих переменных: исходный набор из 10 объясняющих переменных; исходный набор из 10 объясняющих переменных с добавлением 5 переменных, не учитываемых в расчете *Score*; исходный набор из 10 объясняющих переменных, сокращенный до 5 объясняющих переменных; сокращенный до 5 объясняющих переменных при добавлении 5 переменных, не учитываемых в расчете *Score*. Для увеличения выборки до требуемой доли дефолта дублировались наблюдения класса дефолтов в фиксированной области значения одной выбранной объясняющей переменной. В качестве модели классификации использовался алгоритм логистической регрессии в реализации библиотеки Sklearn для языка Python (scikit-learn).

Основные полученные результаты

Для набора данных из 100 наблюдений удаление объясняющих переменных приводит к уменьшению ширины доверительных интервалов метрики $F1$ для всех типов доверительных интервалов. При этом увеличение веса для шума приводит к росту ширины доверительных интервалов для наборов данных, где удалены значимые объясняющие переменные. Для набора данных из 20 000 наблюдений такие эффекты отсутствуют. В табл. 3 указаны средние значения метрик границ доверительных интервалов для различных спецификаций данных.

Таблица 3

Средние значения метрик границ доверительных интервалов

Размер выборки	Признаки	Вес шума	Agresti_Coull	beta	Jeffreys	normal
100	10 + 5	0,1	0,881	0,881	0,881	0,881
		0,5	0,927	0,927	0,927	0,927
		1,0	0,888	0,888	0,888	0,888
	10	0,1	0,857	0,857	0,857	0,857
		0,5	0,895	0,895	0,895	0,895
		1,0	0,860	0,860	0,860	0,860
	5 + 5	0,1	0,510	0,510	0,510	0,510
		0,5	0,706	0,706	0,706	0,706
		1,0	0,673	0,673	0,673	0,673
	5	0,1	0,478	0,478	0,478	0,478
		0,5	0,640	0,640	0,640	0,640
		1,0	0,605	0,605	0,605	0,605
20 000	10 + 5	0,1	0,998	0,998	0,998	0,998
		0,5	0,998	0,998	0,998	0,998
		1,0	0,998	0,998	0,998	0,998
	10	0,1	0,998	0,998	0,998	0,998
		0,5	0,998	0,998	0,998	0,998
		1,0	0,998	0,998	0,998	0,998
	5 + 5	0,1	0,946	0,946	0,946	0,946
		0,5	0,947	0,947	0,947	0,947
		1,0	0,948	0,948	0,948	0,948
	5	0,1	0,944	0,944	0,944	0,944
		0,5	0,945	0,945	0,945	0,945
		1,0	0,945	0,945	0,945	0,945

Для набора данных из 20 000 наблюдений синтетическое увеличение доли дефолтов в выборке приводит к ухудшению метрики $F1$, однако метрика ухудшается максимум на 18% от значения на исходной выборке. В случае набора данных со 100 наблюдениями это падение может достигать 50%. Таким образом, при критично малом количестве наблюдений не рекомендуется синтетически увеличивать долю дефолтов в выборке. Также отмечается, что ширина доверительного интервала для метрики $F1$ увеличивается с добавлением синтетических данных. В табл. 4 указаны средние значения метрики $F1$ для набора данных из 5 объясняющих признаков при размере выборки 20 000. В табл. 5 указана аналогичная таблица для размера выборки в 100 наблюдений.

Для всех наборов данных и вариаций параметров передискретизации различия в типах случайных пропусков и их влияние на метрики качества алгоритмов классификации незначительны или отсутствуют.

Таблица 4

Средние значения метрик границ доверительных для выборки в 20 000 наблюдений

*F*1	До увеличения	После увеличения							Не увеличено
Пропуски		DR = 0,5	DR 1	DR 3	DR 5	DR 10	DR 25	DR 50	
Нет	DR 50								0,755
MCAR	DR 10						0,888	0,757	0,953
	DR 5					0,951	0,887	0,763	0,975
	DR 3				0,977	0,955	0,895	0,772	0,986
	DR 1			0,985	0,976	0,950	0,893	0,801	0,995
	DR 0,5		0,995	0,985	0,978	0,954	0,890	0,820	0,998
	DR 0,1	0,998	0,995	0,985	0,977	0,954	0,888	0,821	0,998
MAR	DR 10						0,881	0,747	0,951
	DR 5					0,950	0,883	0,747	0,975
	DR 3				0,975	0,952	0,884	0,770	0,985
	DR 1			0,985	0,974	0,947	0,865	0,718	0,995
	DR 0,5		0,995	0,985	0,974	0,952	0,889	0,759	0,998
	DR 0,1	0,997	0,995	0,988	0,981	0,960	0,904	0,792	1,000

Таблица 5

Средние значения метрик границ доверительных для выборки в 100 наблюдений

*F*1	До увеличения	После увеличения							Не увеличено
Пропуски		DR = 0,5	DR 1	DR 3	DR 5	DR 10	DR 25	DR 50	
Отсутствуют	DR 50								0,848
MCAR	DR 10						0,529	0,462	0,848
	DR 5					0,551	0,529	0,462	0,848
	DR 3				0,642	0,551	0,529	0,462	0,848
	DR 1			0,682	0,611	0,529	0,529	0,462	0,848
	DR 0,5		0,774	0,675	0,597	0,529	0,529	0,462	0,848
	DR 0,1	0,848	0,739	0,659	0,583	0,529	0,529	0,462	0,848
MAR	DR 10						0,529	0,462	0,848
	DR 5					0,529	0,529	0,462	0,848
	DR 3				0,571	0,529	0,529	0,462	0,848
	DR 1			0,674	0,571	0,529	0,529	0,462	0,848
	DR 0,5		0,739	0,682	0,550	0,551	0,529	0,462	0,848
	DR 0,1	0,825	0,800	0,649	0,649	0,630	0,611	0,529	0,923

Выводы

Проведенный анализ показал значительную зависимость качества моделей кредитного скоринга от объема данных и методов обработки дисбаланса классов. Отмечается, что для малых выборок (100 наблюдений) удаление объясняющих переменных снижает неопределенность (сужение доверительных интервалов *F*1), однако синтетическое увеличение доли дефолтов или добавление шума резко ухудшает метрики (падение $F1$ до 50%) и повышает вариативность оценок. Для крупных выборок (20 000 наблюдений) эти эффекты менее выражены: синтетическое увеличение дефолтов ухудшает $F1$ не более чем на 18%, а устойчивость моделей к шуму и пропускам данных значительно выше. Аналитикам, работающим с низкодефолтными портфелями, рекомендуется избегать синтетической передискретизации. Вместо синтетического увеличения целевого класса целесообразно применять алгоритмы, оптимизированные под редкие события за счет варьирования параметра регуляризации. Для крупных выборок допустимо осторожное использование синтетических методов, но с обязательным мониторингом $F1$ и доверительных интервалов, а также стресс-тестированием на устойчивость к шуму.

Источники

Применение искусственного интеллекта на финансовом рынке: доклад Банка России / Банк России, 2023. URL: https://www.cbr.ru/press/event/?id=17177.

Brown L. D., Cai T. T., DasGupta A. Interval Estimation for a Binomial Proportion // Statistical Science. 2001. Vol. 16. N 2. P. 101–117.

Chawla N. V., Bowyer K. W., Hall L. O., Kegelmeyer W. P. SMOTE: Synthetic Minority Over-sampling Technique // Journal of Artificial Intelligence Research. 2002. Vol. 16. P. 321–357.

Gomer B., Yuan K.-H. A Realistic Evaluation of Methods for Handling Missing Data When There is a Mixture of MCAR, MAR, and MNAR Mechanisms in the Same Dataset // Multivariate Behavioral Research. 2023. Vol. 58. N 5. P. 1–26.

Goorbergh R., Smeden M., Timmerman D., Calster B. The Harm of Class Imbalance Corrections for Risk Prediction Models: Illustration and Simulation Using Logistic Regression // Journal of the American Medical Informatics Association. 2022. Vol. 29. N 9. P. 1525–1534.

Kim J., Jang W. A generalized Agresti–Coull type confidence interval for a binomial proportion // Journal of the Korean Statistical Society. 2022. Vol. 51. N 2. P. 356–377.

LogisticRegression / scikit-learn developers, 2024. URL: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html.

Lott A., Reiter J. P. Wilson Confidence Intervals for Binomial Proportions With Multiple Imputation for Missing Data // The American Statistician. 2020. Vol. 74. N 2. P. 109–115.

Manera M. C. Andrea The Economic Impacts and the Regulation of AI: A Review of the Academic Literature and Policy Actions. Working Paper No. 2024/065. // IMF, 2024. URL: https://www.imf.org/en/Publications/WP/Issues/2024/03/22/The-Economic-Impacts-and-the-Regulation-of-AI-A-Review-of-the-Academic-Literature-and-546645.

Pradeep Reddy G., Rohan D., Kumar Y. V. P., Prakash K. P., Srikanth M. Artificial Intelligence-Based Effective Detection of Parkinson's Disease Using Voice Measurements // Engineering Proceedings. 2024. Vol. 82. N 1. P. 28.

Puza B., O'Neill T. Generalised Clopper–Pearson Confidence Intervals for the Binomial Proportion // Journal of Statistical Computation and Simulation. 2006. Vol. 76. N 6. P. 489–508.

Raschka S., Mirjalili V. Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-Learn and TensorFlow 2. 3rd ed. Packt Publishing, Birmingham, 2019.

Xing Q., Yu C., Huang S., Zheng Q., Mu X., Sun M. Enhanced Credit Score Prediction Using Ensemble Deep Learning Model //arXiv preprint arXiv:2410.00256. 2024.

References

Brown L. D., Cai T. T., DasGupta A. Interval Estimation for a Binomial Proportion. *Statistical Science*, 2001, vol. 16, N 2, pp. 101–117.

Chawla N. V., Bowyer K. W., Hall L. O., Kegelmeyer W. P. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 2002, vol. 16, pp. 321–357.

Gomer B., Yuan K.-H. A Realistic Evaluation of Methods for Handling Missing Data When There is a Mixture of MCAR, MAR, and MNAR Mechanisms in the Same Dataset. *Multivariate Behavioral Research*, 2023, vol. 58, N 5, pp. 1–26.

Goorbergh R., Smeden M., Timmerman D., Calster B. The Harm of Class Imbalance Corrections for Risk Prediction Models: Illustration and Simulation Using Logistic Regression. *Journal of the American Medical Informatics Association*, 2022, vol. 29, N 9, pp. 1525–1534.

Kim J., Jang W. A generalized Agresti–Coull type confidence interval for a binomial proportion. *Journal of the Korean Statistical Society*, 2022, vol. 51, iss. 2, pp. 356–377.

LogisticRegression. scikit-learn developers, 2024. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html.

Lott A., Reiter J. P. Wilson Confidence Intervals for Binomial Proportions With Multiple Imputation for Missing Data. *The American Statistician*, 2020, vol. 74, iss. 2, pp. 109–115.

Manera M. C. *Andrea The Economic Impacts and the Regulation of AI: A Review of the Academic Literature and Policy Actions*. Working Paper No. 2024/065. IMF, 2024. Available at: https://www.imf.org/en/Publications/WP/Issues/2024/03/22/The-Economic-Impacts-and-the-Regulation-of-AI-A-Review-of-the-Academic-Literature-and-546645.

Pradeep Reddy G., Rohan D., Kumar Y. V. P., Prakash K. P., Srikanth M. Artificial Intelligence-Based Effective Detection of Parkinson's Disease Using Voice Measurements. *Engineering Proceedings*, 2024, vol. 82, N 1, pp. 28.

Primeneniye iskusstvennogo intellekta na finansovom rynke: doklad Banka Rossii [Application of Artificial Intelligence in the Financial Market: Bank of Russia Report]. Bank Rossii [Bank of Russia], 2023. Available at: https://www.cbr.ru/press/event/?id=17177. (In Russian)

Puza B., O'Neill T. Generalised Clopper–Pearson confidence intervals for the binomial proportion. *Journal of Statistical Computation and Simulation*, 2006, vol. 76, N 6, pp. 489–508.

Raschka S., Mirjalili V. *Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-Learn and TensorFlow 2*. 3rd ed. Packt Publishing, Birmingham, 2019.

Xing Q., Yu C., Huang S., Zheng Q., Mu X., Sun M. Enhanced Credit Score Prediction Using Ensemble Deep Learning Model. *arXiv preprint arXiv:2410.00256*, 2024.

Статья поступила в редакцию 19 февраля 2025 г.
Статья рекомендована в печать 15 июня 2025 г.