DOI 10.15826/vestnik.2025.24.4.045 Original Paper

УДК 336.67

# Отбор информативных показателей для оценки экономической безопасности российских компаний

*Л. А. Буланов*[1], *А. В. Калина*[1, 2] ✉, *В. В. Криворотов*[1]

[1] *Уральский федеральный университет имени первого Президента России Б. Н. Ельцина, г. Екатеринбург, Россия*

[2] *Институт экономики Уральского отделения РАН, г. Екатеринбург, Россия*

✉ *alexkalina74@yandex.ru*

***Аннотация.*** Задача обеспечения экономической безопасности предприятий и компаний является одной из первоочередных в современных экономических и политических условиях. Центральное место в решении этой задачи является всесторонняя объективная оценка экономической безопасности компаний, в основе которой, как правило, лежит метод индикативного анализа. Это, в свою очередь, требует разработки и формирования состава индикаторов экономической безопасности, которое бы наиболее полно и качественно отражали состояние компаний как в текущий, так и перспективный периоды. Целью настоящей статьи является разработка алгоритма и его модельная реализация для отбора наиболее значимых показателей экономической безопасности предприятий и компаний, основанного на использовании технологий машинного обучения. Для реализации поставленной цели в работе исследуется интерпретируемая многоклассовая классификация крупных производственных компаний («плохие/нормальные/хорошие») по их финансовым коэффициентам с формированием целевых меток кластеризацией K-Means++ на реальном датасете за 2023 г. (2 249 наблюдений). Сопоставляются одномерные фильтры (ANOVA/$\eta^2$, one-vs-rest ROC-AUC, Mutual Information) и модельные важности SHAP для CatBoost; сравнение проводится через корреляцию рангов, пересечение топ-k и абляции с логлосс-оценкой на стратифицированной k-fold CV при строгом разделении train/val. Дополнительно оценивается компромисс «компактность vs точность» при $k \in \{5, 10, 15, 20, 30\}$ и проверяется статистическая значимость различий (критерий Уилкоксона). Предложенный подход и его реализация для российских компаний выявили, что собственные оборотные средства являются основой для формирования и оценки наиболее значимых и информативных индикаторов экономической безопасности компании. При этом профили по классам показывают асимметричные, экономически согласованные границы: для «плохих» компаний решающим является дефицит оборотного капитала и низкая ликвидность; для «хороших» компаний — высокие быстрая/текущая ликвидность и покрытие; «нормальные» компании формируются как буфер при умеренных значениях и обслуживаемом долге. В качестве главного практического результата проведенного исследования можно выделить практические рекомендации по минимальному набору показателей и стратегии отбора признаков для мониторинга экономической безопасности и финансового состояния компаний и прикладной аналитики в этой области.

***Ключевые слова:*** показатели экономической безопасности; машинное обучение; кросс-валидация; отбор признаков; ANOVA; ROC-AUC; Mutual Information; CatBoost; SHAP.

## 1. Введение

Оценка и обеспечение экономической безопасности (ЭБ) предприятий и компаний являются одной из первоочередных задач, стоящих перед ними. Прежде всего такая оценка лежит в основе принятия стратегических и текущих решений по развитию компаний, поддержания нормальных условий их деятельности, противодействия угрозам и кризисным явлениям, то есть для решения широкого круга задач, связанных с текущим функционированием и перспективным развитием компаний. Это, в свою очередь, требует разработки методики комплексной оценки экономической безопасности предприятий и компаний, основанной на использовании широкого круга показателей и индикаторов, которые оценивают разные стороны деятельности и направления обеспечения ЭБ компаний.

Вместе с тем универсальных общепринятых подходов к формированию состава индикативных показателей и проведению оценки ЭБ предприятий и компаний на сегодняшний день нет. Многие ученые и специалисты отдают приоритет тем или иным сторонам и аспектам деятельности компаний (финансы, экономика, производство, маркетинг и др.), не учитывая или учитывая очень укрупнено другие. При этом во многих случаях при формировании состава индикаторов присутствует взаимодублирование, слабая информационная обеспеченность, малая информативность и множественное толкование отдельных показателей и др., что, с одной стороны, усложняет решаемую задачу, а с другой — существенно снижается ценность и значимость получаемых результатов. Это, в свою очередь, делает очень важным решение вопросов и проблем, связанных с формированием объективно значимой универсальной системы индикаторов экономической безопасности компаний, учитывающей различные аспекты их деятельности.

В таких условиях наиболее рациональным подходом к формированию индикаторов ЭБ предприятий и компаний и последующей ее оценке является изначальный отбор наиболее значимых и информативных показателей ЭБ и их последующее использование для оценки и анализа. Для решения данной задачи предлагается использовать современные подходы, основанные на применение технологий машинного обучения, ранее успешно использовавшиеся авторами при решении подобных задач [1].

Как следствие, одним из наиболее важных вопросов проводимого исследования является обоснование и выбор методов и подходов к отбору информативных показателей и классификации компаний по экономической безопасности. Кроме того, важным вопросом при использовании нескольких методов и подходов является разработка рекомендаций по практическому использованию и ограничениям для использования отдельных методов, а также выбору наиболее рационального метода, исходя из конкретных условий и ограничений.

Говоря о современных подходах к отбору наиболее информативных признаков, можно отметить, что интерпретируемость и осознанный отбор

признаков в финансовой аналитике выступают как неотъемлемые условия формирования эффективной стратегии, поскольку от прозрачности моделей зависят управленческие решения, регуляторная приемлемость и устойчивость выводов. Классическая таксономия методов отбора признаков — фильтры, обертки и встроенные методы — различается степенью связности с конечной моделью и вычислительной стоимостью. Одномерные фильтры (ANOVA/$\eta^2$, Kruskal-Wallis/$\varepsilon^2$, one-vs-rest ROC-AUC, взаимная информация) просты и интерпретируемы, но по построению измеряют лишь маргинальную связь признака с целью, игнорируя взаимодействия и конкурирующие переменные.

Современные модельно-агностичные подходы интерпретации, такие как SHAP/TreeSHAP, опираются на аддитивные атрибуции Шэпли и согласованно раскладывают предсказание на вклады признаков, предоставляя локальные и глобальные важности. Вместе с тем в присутствии скоррелированных признаков распределение важности между переменными становится неоднозначным и требует осторожной интерпретации. В условиях стремительной цифровизации и роста данных особенно актуальны методы, одновременно обеспечивающие прозрачность, воспроизводимость и операционную пригодность решений.

В настоящем исследовании рассматривается задача трехклассовой классификации производственных компаний («плохие/нормальные/хорошие») по публичной отчетности и выбор на этой основе наиболее информативных показателей, где целевые классы заданы кластеризацией в пространстве финансовых показателей. Такая постановка естественна для экономической интерпретации: именно коэффициенты ликвидности, левереджа, маржинальности и эффективности исторически демонстрируют прогностическую силу в задачах риск-сегментации и предсказания несостоятельности (банкротства). Классификация строится на основе сопоставления одномерных метрик важности (фильтров) с модельными важностями (SHAP) на реальном многоклассовом финансовом датасете за один год и на количественной оценке потерь качества при вынужденном сокращении числа признаков.

*Целью* исследования является разработка алгоритма и его модельная реализация для отбора наиболее значимых показателей экономической безопасности предприятий и компаний, основанного на использовании технологий машинного обучения.

В работе рассматриваются следующие гипотезы:

*H1*: С использованием современных технологий машинного обучения возможно провести объективное ранжирование показателей экономической безопасности и финансового состояния предприятий и компаний, выделив при этом наиболее информативные показатели, имеющие наибольшую важность для оценки ЭБ.

*H2:* Ранжирования признаков, полученные одномерными фильтрами (ANOVA/$\eta^2$, one-vs-rest ROC-AUC, взаимная информация), будут лишь частично согласованы с ранжированием по SHAP для многоклассовой модели

градиентного бустинга по деревьям (CatBoost), поскольку взаимодействия и конкуренция признаков меняют их условную важность.

*H3:* При сокращении признакового пространства качество классификации будет деградировать медленнее для подсетов, отобранных по глобальным важностям SHAP, чем для подсетов, выбранных одномерными фильтрами, что отражает преимущество учета совместного влияния признаков.

*H4:* Сильная коррелированность признаков приведет к перераспределению важности и снижению устойчивости ранжирований в повторных выборках, причем эффект будет более выраженным у одномерных фильтров, тогда как SHAP важности сохранят информативность, но потребуют осторожной интерпретации ввиду деления вклада между зависимыми переменными.

*Структура статьи*. Во введении обоснована актуальность проводимого исследования, приведены цель и гипотезы проводимого исследования. Во втором разделе дан аналитический обзор научной литературы, связанной с формированием подходов и показателей для оценки ЭБ компаний, а также представлен анализ литературы по отбору наиболее информативных признаков для классификации показателей и компаний. В третьем разделе приведена методология исследования и дана характеристика данных, на основе которых проводилось исследование. В четвертом разделе показаны и обоснованы основные результаты проведенного исследования. В пятом разделе проводится обсуждение полученных результатов и даются рекомендации по их использованию. В шестом разделе приведены основные выводы по результатам проведенного исследования.

## 2. Обзор литературы

### *2.1. Подходы и оценка экономической безопасности и ее составляющих*

Исследования экономической безопасности в последние десятилетия привлекают большое внимание со стороны многих специалистов и ученых, занимающихся экономическими исследованиями и решающими практические задачи в этой области. Соответственно накоплен богатый опыт исследования для объектов разного уровня и иерархии. При этом в первую очередь в качестве объектов исследования и обеспечения экономической безопасности рассматриваются территории (государство, регион), предприятия и компании, личность.

Первоначально главным объектом исследования экономической безопасности выступало государство[1]. Соответственно, исследования для таких объектов, в первую очередь, были ориентированы на создании системы защиты национальных экономик стран от внутренних и внешних угроз и построении

[1] Понятие экономической безопасности было введено в оборот в 1930-е гг. в США в период развития и преодоления последствий крупнейшего экономического кризиса — Великой депрессии. Подход к пониманию национальной экономической безопасности был сформулирован английским ученым Джоном Мейнардом Кейнсом.

на этой основе стратегии долгосрочного экономического развития страны и входящих в нее территорий (регионы и т. п.).

В качестве одного из современных примеров таких работ можно выделить статью Hauge et al. [2], где рассматриваются вопросы реализации промышленной политики ведущих стран (США, Китай, Европейский Союз) в направлении достижения приоритетов национальной экономической безопасности. При этом в качестве одной из главных тенденций последних лет авторы выделяют усиление экономического национализма как одного из главных подходов в обеспечении ЭБ страны.

Во многом аналогичные взгляды отражены в работе Ahumada [3], где обеспечение экономической безопасности страны в современном мире видится в индустриализации и национальной независимости.

Несмотря на важность и значимость обеспечения экономической безопасности на уровне государства, работ зарубежных ученых, которые развивают эту проблему относительно немного и термин «экономическая безопасность» встречается достаточно редко. Большая часть работ по проблеме экономической безопасности посвящена развитию ее отдельных составляющих, в первую очередь энергетической, продовольственной и экологической (эколого-экономической) безопасности, что определяются высокой актуальностью и важностью этих проблем в современном мире.

Прежде всего остановимся на энергетической безопасности. Ghazouani [4] исследует влияние геополитических, технологических и экономических факторов на энергетическую безопасность страны на примере США с применением процедур квантильно-вейвлет-анализа. Исследование показало прямое влияние указанных факторов на энергетическую безопасность и позволило сформировать практические решения, позволяющие сбалансировать энергетическую безопасность, экономический рост и экологическую устойчивость.

Во многом схожие исследования энергетической безопасности на уровне государства проводились Akusta [5]. Им анализировалось влияние глобализации на риски с позиций обеспечения энергетической безопасности для 26 стран ОЭСР, используя данные за период 1980–2018 гг. Результаты показали, что экономическая, политическая и совокупная глобализация увеличивает риск энергетической безопасности, в то время как социальная глобализация статистически незначима. Также автор выявил, что влияние глобализации на риски энергетической безопасности различается в разных странах. Это, соответственно, приводит к тому, что политика минимизации рисков и обеспечения энергетической безопасности в различных странах должна отличаться.

Многие исследования посвящены формированию системы показателей и оценке энергетической безопасности в большинстве случаев на уровне государства. Так, Yao & Chang [6] проводили количественную оценку энергетической безопасности Китая. В основу формирования показателей для оценки были положены четыре основные направления: наличие энергетических

ресурсов, применимость технологий, приемлемость для общества и доступность энергетических ресурсов по цене.

Работы Kruyt et al. [7] и Löschel [8] связаны с формированием системы показателей (индикаторов) энергетической безопасности государства, что является основной для проведения ее оценки. Авторы достаточно глубоко проработали этот вопрос и предложили достаточно полные и детальные системы показателей, позволяющих провести глубокую оценку и анализ энергетической безопасности и выработать на этой практические рекомендации по ее обеспечению и минимизации рисков. Однако речи о создании унифицированной универсальной системы показателей энергетической безопасности здесь не идет, каждый из авторских коллективов придерживается своего видения по этому вопросу, хотя во многом их взгляды пересекаются.

Одной из наиболее значимых в направлении формирования системы показателей и индексов для оценки энергетической безопасности является статья Ang et al. [9], где авторы на основе более 100 исследований по проблеме предложили свою систему показателей для оценки энергетической безопасности страны, которая полноценно может претендовать на унифицированную универсальную систему и быть внедрена в практическую деятельность органов власти и компаний большинства стран.

Высокую значимость занимает проблема продовольственной безопасности. В первую очередь, в данном направлении развиваются вопросы создания устойчивых производственных продовольственных систем и обеспечения населения доступными и качественными продуктами питания. Так, Abd El-Ghani et al. [10] рассматривают основные экономические, экологические и социальные проблемы, влияющие на продовольственную безопасность на примере Египта. Вопросы и задачи построения устойчивых продовольственных систем исследуются и анализируются в статье Akinci1 & Kumcu [11], а примером подхода к решению проблемы обеспечения продовольственной безопасности в условиях глобализации и роста численности населения в неразвитых странах является работа Onwe et al. [12], где эти вопросы исследуются на примере Нигерии.

Переходя к оценке экономической безопасности на уровне предприятий и компаний следует отметить следующее. В трудах зарубежных ученых термин «экономическая безопасность» (“economic security”) в чистом виде встречается редко. Как правило, исследования в этой области посвящены проблеме риск-менеджмента, оценке отдельных составляющих экономической безопасности компаний (кадровой, информационной, финансовой и др.), а также организации и работе служб компаний, ответственных за обеспечение экономической безопасности и ее составляющих.

Наиболее близкими к тематике оценки экономической безопасности предприятий и компаний являются исследования, связанные с риск-менеджментом в компаниях. В последние годы наибольшее развитие в этом направлении получили работы, связанные с управлением ESG-рисками и управлением

рисками, связанным с устойчивым развитием. Так, например, Boiral et al. [13] рассматривают вопросы управления ESG-рисками в горнодобывающих компаниях. Авторы на основе 57 опасных ситуаций в 19 горнодобывающих компаниях, что способствовало повышению предсказуемости экологических, социальных и управленческих рисков на начальной стадии разработки проектов по добыче полезных ископаемых.

Bertolotti [14] рассматривает вопросы интеграции ESG-факторов в дополнение к традиционным финансовым рискам в портфели инвестиционных проектов, реализуемых компаниями. Автор предлагает характеристики портфеля, связанные с интеграцией данных ESG, а также подходы и проблемы, связанные с созданием инструментов для измерения рисков.

Примером реализации подходов к управлению рисками, связанным с устойчивым развитием является работа Razak et al. [15], где авторы провели объемное исследование среди малазийских компаний и оценили их готовность к внедрению системы управления рисками устойчивого развития (SRM).

Обеспечение экономической и финансовой безопасности в контексте управления кредитными рисками рассматривается в работе Belás et al. [16]. Кредитный риск и платежеспособность как основа оценки и обеспечения экономической безопасности компаний встречаются в работах многих авторов. Среди примеров можно привести труды Altman et al. [17], Behr & Guttler [18] и многих других специалистов.

Вопросы формирования методического аппарата и проведения непосредственной оценки экономической безопасности предприятий и компаний чаще встречаются в трудах ученых из стран восточной Европы. Среди наиболее известных в этом направлении следует выделить работу Pylypenko et al. [19], в которой были выделены ключевые компоненты системы экономической безопасности предприятий и компаний и предложена структурная модель ее обеспечения.

Одним из наиболее качественных подходов оценке экономической безопасности компаний являются работы Ianioglo [20, 21], где предложен комплексный методический инструментарий оценки ЭБ предприятий и компаний, где автор постарался учесть различные составляющие и направления деятельности компаний.

Следует также выделить работу Ereshko & Karanina [22], в которой авторами предлагается и обосновывается подход, основанный на использовании факторов ускоряющих (замедляющих) расчетные показатели количественных значений экономической безопасности. При этом предлагается система показателей многокритериальной оценки экономической безопасности применительно к транспортным предприятиям.

Методика оценки экономической безопасности автотранспортных предприятий предложена Ryazanova & Timin [23]. Авторы предлагают использовать четыре группы показателей для оценки экономической безопасности таких предприятий, а именно: качество дорожной сети, эффективность

использования автомобильного транспорта, эффективность использования автобусного транспорта, эффективность транспортных предприятий. В то же время следует сказать, что такой подход заточен в основном под автотранспортные предприятия и малоприменим для предприятий и компаний, работающих в других сферах деятельности.

Prievozník et al. [24] для оценки экономической безопасности компании предлагают использовать трехмерную модель, которая реализуется на примере поставщиков услуг общественного транспорта. В своем подходе авторы используют показатели годовых отчетов поставщиков услуг общественного транспорта в сочетании с методологией диаграмм Исикавы.

Подытоживая анализ подходов к оценке экономической безопасности предприятий и компаний, можно сказать, что в настоящее время общепринятого универсального подхода к такой оценке нет. Его разработка является значимой актуальной задачей. Одним из центральных мест в подходе должна занимать системы показателей (индикаторов) экономической безопасности компании. В основу ее формирования предлагается положить процедуры отбора наиболее информативных показателей, которые предлагаются в настоящем исследовании. Система показателей экономической безопасности предприятий и компаний, сформированная с использованием такого подхода, должна обладать высокими диагностическими свойствами и характеристиками при проведении непосредственной оценки ЭБ предприятий и компаний.

### *2.2. Подходы к отбору наиболее значимых признаков и информативных показателей*

Методы и подходы к отбору наиболее информативных показателей и признаков в отечественной и зарубежной науке развиваются достаточно давно. Изначально в основу подходов был положен дискриминантный анализ и широкое использование методов экспертного анализа. Одним из наиболее известных ранних подходов является работа Beaver [25], в которой автор применил унивариантный анализ — сравнение распределений отдельных финансовых коэффициентов у «здоровых» и «проблемных» фирм и их ранжирование по дискриминационной способности.

Altman [26] в основу отбора положил пошаговый дискриминантный анализ для выбора подмножества коэффициентов, максимизирующих разделение классов. Итоговый набор состоял из пяти показателей (Z-score).

Ohlson [27] использовал логистическую регрессия. При этом значимость признаков оценивалась через статистики Wald и Likelihood Ratio, а информативность через прирост лог-правдоподобия. Следует также отметить, что ограниченные возможности средств вычислительной техники в XX столетии не позволяли многим ученым и специалистам использовать многие современные продвинутые методы ввиду невозможности или ограниченной возможности их практической реализации.

В XXI столетии возможности по реализации различных методов к отбору информативных признаков и показателей существенно выросли, что дало скачок в реализации новых подходов к решению этой задачи. Shumway [28] применил панельную событийную (hazard) модель. Эта модель использовала как стандартные показатели бухгалтерской отчетности, так и рыночные переменные для составления прогнозов на основе выборки, которые являются более точными, чем прогнозы альтернативных моделей. При этом автор выявил, что многие показатели финансовой отчетности, которые использовались ранее, во многих исследованиях не являются статистически значимыми.

Сравнение современных алгоритмов (логит, деревья, SVM, бустинг) с фильтрами/врапперами отбора признаков и встроенной регуляризацией лежит в основе подхода, предложенного Lessmann et al. [29]. В первую очередь реализация и развитие этого подхода были направлены на выделение и сравнение информативных признаков для оценки кредитоспособности компаний.

Saisana et al. [30] в основу отбора информативных показателей и формирования состава показателей для страновых сравнений и рейтинговой оценки отдельных стран предлагают использовать анализ неопределенности и чувствительности. В качестве результирующего показателя в работе используется индекс технологических достижений, состоящий из широкого набора частных показателей, значимость которых оценивается на основе разложения дисперсии индекса по влиянию этих показателей.

Развитие модели Z-Score, предложенной Altman, рассматривается в работе [31], где предлагается перекалибровка и адаптация набора показателей Z-score для разных стран/рынков и сравнение релевантности значимости признаков по юрисдикциям для решения задачи прогнозирования банкротства и других проблемных ситуаций, которые возникают в деятельности компаний. На основании исследования компаний более чем 30 стран авторы пришли к выводу, что модернизированная модель Z-Score достаточно хорошо работает для большинства стран (точность прогноза составляет приблизительно 0,75), а точность классификации может быть повышена еще больше (выше 0,90) за счет использования для оценки дополнительных переменных специфичных для каждой конкретной страны.

В настоящее время в методах отбора признаков традиционно выделяют три класса подходов — фильтровые, обертки и встроенные (embedded). Они по-разному балансируют вычислительную стоимость, устойчивость и связь с конечной моделью. Фундаментальные обзоры подчеркивают, что фильтры полезны как быстрый и интерпретируемый скрининг, но по определению оценивают лишь маргинальную (одномерную) связь признака с целью и игнорируют взаимодействия и конкуренцию признаков [32, 33], что подтверждается и современными обзорами последних лет [34].

В качестве одномерных фильтров в многоклассовых задачах широко применяются дисперсионный анализ ANOVA (с интерпретацией через эффект-размер $\eta^2$) и непараметрический критерий Краскела — Уоллиса (Kruskal &

Wallis) эффект-размер $\varepsilon^2$ [35], а также ROC-методология: одномерная «разделяющая способность» оценивается через AUC в постановке one-vs-rest с инвариантностью к направлению (берется max(AUC, 1–AUC)) и корректной агрегацией на многоклассовый случай [36, 37]. В частности, с помощью ANOVA Bademlioglu et al. [38] оценивали вклад параметров радиатора в эффективность охлаждения, а Aravind Kumar et al. [39] использовали критерий Краскела — Уоллиса для оценки модели растворимости антираковых препаратов.

Для фиксации нелинейной зависимости и работы со смешанными типами данных используется взаимная информация (MI), в частности KNN оцениватели и их модификации для непрерывно дискретных признаков [40, 41]. Например, в статье Zhang et al. [42] MI используется в построении модели движения судна при маневрах. При множественных проверках стандартом контроля ложных открытий является FDR-подход Бенджамини — Хохберга (BH). Так, в современных пайплайнах одно клеточного анализа (Seurat) при выявлении дифференциальной экспрессии и интеграции мультиомных данных систематически контролируют FDR по BH [43].

На стороне модельной интерпретации де-факто стандартом стали аддитивные атрибуции Шэпли — SHAP/TreeSHAP, обеспечивающие согласованное (в смысле аксиом Шэпли) разложение предсказания на вклады признаков, с локальными объяснениями на уровне объекта и глобальными важностями (например, через среднее абсолютное значение SHAP) [44, 45]. Метод активно применяется в прикладных областях — от медицины (прогноз гипоксемии в реальном времени) [46] и до сферы финансов (построение кредитного скоринга) [47].

В научной литературе подчеркиваются преимущества SHAP над сугубо «внутримодельными» важностями деревьев по gain/split, которые подвержены смещению, особенно при различиях в масштабе и при наличии категориальных признаков, но одновременно дается предупреждение о неоднозначности атрибуций при зависимых (скоррелированных) признаках и рекомендуется осторожная интерпретация либо группировка коррелирующих переменных [48, 49]. В качестве независимой проверки часто используют пермутационную важность, идущую от работ по случайным лесам и затем обобщенную. Она широко применяется вместе с SHAP для валидации важностей и выявления эффектов корреляции в современных прикладных исследованиях.

В частности, Oladimeji et al. [50] используют комбинацию пермутационной важности и SHAP для диагностики рака молочной железы. Zhou et al. [51] аналогично используют оба этих метода для оценки взаимосвязи между различными факторами энергопотребления в офисных зданиях и выбросами углекислого газа.

В качестве базовой модели для эмпирической части применяется градиентный бустинг над решающими деревьям [52]. Наряду с широко

используемыми XGBoost [53] и LightGBM [54], в задачах со смешанными типами признаков и категорий особое место занимает CatBoost, использующий упорядоченное бустирование и специальное кодирование категориальных признаков, что уменьшает переобучение и снижает чувствительность к тюнингу гиперпараметров [55]. В настоящее время градиентный бустинг очень активно применяется в различных сферах деятельности, например Zhang et al. [56] оценивают прочность бетона на сжатие с использованием CatBoost. Arif et al. [57] при помощи LightGBM ищут противораковые пептиды.

На фоне пары «фильтры против SHAP» важно контекстуализировать альтернативные стратегии отбора: RFE (включая SVM RFE) как класс обертоk [58], последовательные прямые и «плавающие» методы [59, 60], Boruta/BorutaSHAP как сравнение с «теневыми» признаками [61], а также встроенные L1/LASSO [62] и Elastic Net [63] с их регуляризационными свойствами и stability selection как способ повысить устойчивость выбора к случайным флуктуациям данных [64].

Ключевым для валидности эмпирического сравнения является корректный дизайн валидации. Показано, что, если отбор признаков или настройка гиперпараметров использует информацию из валидации, оценка качества смещается в оптимистичную сторону. Отсюда отбор делается строго на train-выборке и при необходимости вложенной кросс валидации [65, 66]. Для статистического сравнения стратегий на зависимых по фолдам выборках уместны непараметрические парные тесты (например, критерий Уилкоксона) [67].

### *2.3. Формирование исследовательских разрывов*

Несмотря на широкое применение одномерных фильтров в многоклассовых задачах (дисперсионного анализа с оценкой эффекта через $\eta^2$, критерия Краскела – Уоллиса с $\varepsilon^2$, а также OvR-ROC-AUC и взаимной информации, в том числе KNN-оценивателей для смешанных типов), существующие подходы в основном фиксируют лишь одномерные связи между признаком и целевой переменной. Такой фокус на одиночных предикторах игнорирует взаимодействия и конкуренцию признаков, что особенно критично для финансово-экономических данных с богатой зависимой структурой.

Дополнительно MI хотя и улавливает нелинейности, остается одномерной и чувствительной к выбору параметров (дискретизация, число соседей) и объему выборки. OvR-агрегация AUC для многоклассовых задач варьирует по схемам усреднения, что затрудняет сопоставимость результатов. В результате нет согласованного и воспроизводимого ранжирования индикаторов экономической безопасности, подтвержденного вне конкретных выборок и методических настроек.

Также отсутствуют систематические сравнения одномерных фильтров с модельно-зависимыми важностями, способными учитывать совместное влияние признаков, прежде всего с глобальными важностями на основе SHAP

для древовидных ансамблей в многоклассовой постановке. Наличие работ по permutation importance или бинарным задачам не закрывает вопрос: мало исследований, где прямо сопоставляются ранги, оценивается их согласованность (ранговые корреляции, пересечение топ-k) и устойчивость на повторных разбивках и/или out-of-time сценариях. Для предметной области экономической безопасности, где существенны взаимодействия (например, ликвидность и долговая нагрузка, структура капитала и волатильность потоков), такая сравнительная перспектива почти не представлена, а существующие рейтинги индикаторов часто имеют экспертно-описательный характер без проверки прогностической валидности.

Еще один разрыв связан с тем, что в литературе редко оценивается прогностическая валидность ранжирования через «кривые деградации» качества при последовательном сокращении признакового пространства. Часто ограничиваются статистической значимостью отдельных тестов, не показывая, как выбор топ-k признаков по тому или иному критерию сказывается на итоговой многоклассовой классификации (macro/micro-F1, balanced accuracy, logloss, macro-ROC-AUC). Без такой проверки невозможно говорить об объективности важностей: необходимо демонстрировать стабильность и практическую полезность ранжирования в downstream-моделях.

Наконец, влияние мультиколлинеарности на распределение важности и устойчивость ранжирований изучено недостаточно. В экономических данных многие индикаторы тесно коррелированы, что ведет к дублированию сигналов и размазыванию вкладов. Одномерные фильтры могут переоценивать избыточные признаки и давать нестабильные топ-списки при небольших изменениях выборки. В модельно-зависимых методах (включая SHAP) важность перераспределяется внутри кластеров зависимых переменных, требуя кластерной интерпретации. Однако практики такой интерпретации и ее эмпирические эффекты в многоклассовых задачах почти не описаны. Ситуацию усугубляет то, что в скрининге признаков не всегда контролируются множественные проверки (например, через FDR-процедуры), что повышает риск ложноположительных выводов о значимости отдельных индикаторов.

В совокупности эти пробелы — отсутствие стабильных и валидированных ранжирований для индикаторов экономической безопасности, недостаток прямых сравнений одномерных фильтров с SHAP в многоклассовых ансамблях, отсутствие оценки качества по кривым деградации, а также непроработанность эффектов коррелированности на устойчивость и интерпретацию важностей — и формируют основу для настоящего исследования. Они мотивируют проверку того, обеспечивают ли SHAP-важности более стабильное и прогностически валидное ранжирование, в какой мере их ранги согласуются с одномерными фильтрами, как разные критерии влияют на деградацию качества при отборе признаков и как коррелированность признаков перераспределяет важность и снижает устойчивость ранжирования.

## 3. Данные и методы

### *3.1. Характеристика набора данных для исследования*

Рабочий набор данных, использованных при исследовании, после очистки состоит из 2 249 наблюдений, которые представляют из себя отчетность крупных (имеющих доход более 2 миллиардов рублей в год и штатом от 251 сотрудника) производственных (относящихся к производственным группам ОКВЭД) компаний за 2023 г. Для каждой компании были рассчитаны показатели финансовой устойчивости, ликвидности, оборачиваемости, а также финансовых результатов. Все посчитанные характеристики по группам представлены в табл. 1. Кластеризацией методом K-Means++ [1] были получены метки, оценивающие компании как «плохие», «нормальные» и «хорошие», итоговый баланс классов составил 53, 29 и 18 % соответственно.

Таблица 1. **Финансовые показатели компаний, учитываемые при исследовании**
Table 1. **Financial indicators of companies taken into account in the study**

| Показатели финансовой устойчивости | Показатели оборачиваемости | Показатели финансовых результатов | Показатели ликвидности |
|---|---|---|---|
| Коэффициент автономии | Оборачиваемость активов | Рентабельность активов | Коэффициент абсолютной ликвидности |
| Коэффициент капитализации | Оборачиваемость дебиторской задолженности | Рентабельность по чистой прибыли | Коэффициент быстрой ликвидности |
| Коэффициент обеспеченности запасов | Оборачиваемость запасов | Рентабельность продаж | Коэффициент текущей ликвидности |
| Коэффициент покрытия активов | Оборачиваемость кредиторской задолженности | Рентабельность задействованного капитала | Коэффициент обеспеченности собственными средствами |
| Коэффициент покрытия инвестиций | Оборачиваемость оборотных средств | Рентабельность собственного капитала | |
| Коэффициент покрытия процентов | Фондоемкость | | |
| Коэффициент финансовой зависимости | Фондоотдача | | |
| Коэффициент финансового левериджа | | | |

*Источник*: составлено авторами.

Распределение компаний по ОКВЭД представлено на рис. 1. Заметно, что порядка трети компаний — это пищевое производство. Остальные группы значительно меньше, в частности следующая группа — это производство минеральной продукции с долей порядка 10 %. Интересно, что распределение компаний в разрезе видов деятельности при анализе общих активов меняется не сильно. Доля первой группы снизилась примерно на 5 %, доля остальных чуть возросла (рис. 2).

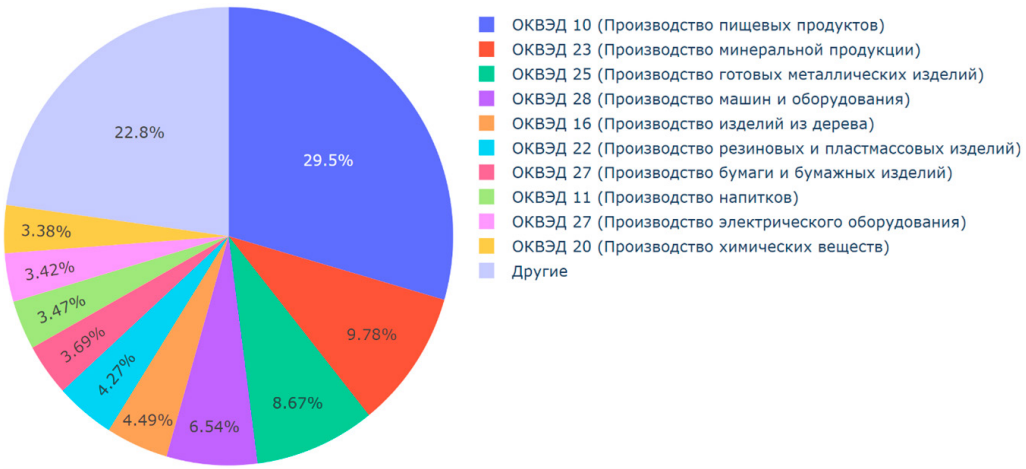


**Рис. 1.** Диаграмма распределения компаний по ОКВЭД

**Fig. 1.** Diagram of the distribution of companies by OKVED (Russian National Classifier of Types of Economic Activity)

*Источник*: составлено авторами.

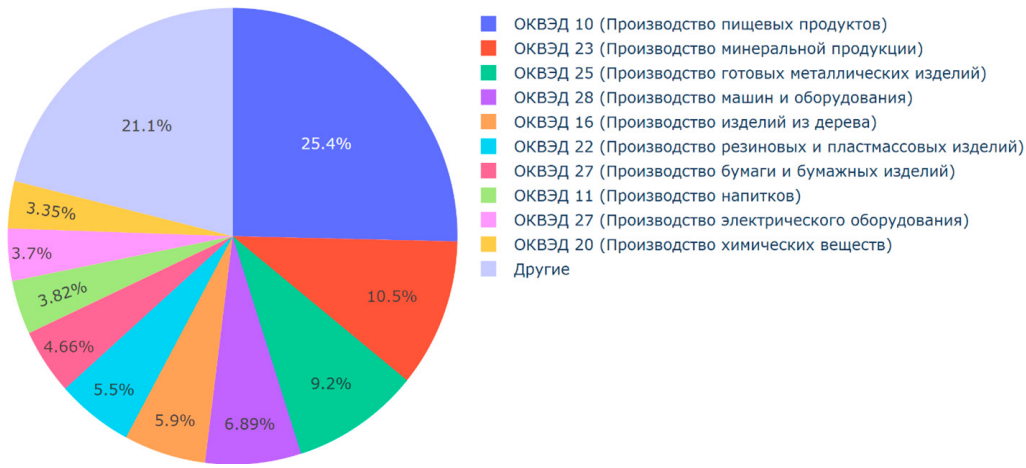


**Рис. 2.** Диаграмма распределения активов компаний по ОКВЭД

**Fig. 2.** The diagram of the distribution of assets of companies by OKVED (Russian National Classifier of Types of Economic Activity)

*Источник:* составлено авторами.

Для оценки однородности данных рассмотрим «ящик с усами» по активам (рис. 3). Заметим, что, для того чтобы сделать его несколько нагляднее, верхняя граница диапазона активов была несколько снижена, поэтому порядка 10 самых больших компаний график не показывает. Тройка крупнейших компаний по активам в датасете представлена в табл. 2.

В целом в датасете оказалось порядка 8,5 % выбросов, что не слишком много, но дисперсия при этом относительно высокая, большая часть компаний расположены в промежутке от примерно одного до пяти миллиардов рублей активов. Максимумы в лице гигантов, представленных в табл. 1, находятся очень далеко от медианных значений. Однако в данном исследовании это не проблема, поскольку использовались не балансовые данные, а посчитанные на их основе коэффициенты. Фильтрация выбросов происходила именно на них.

Для идентификации выбросов использовался метод, основанный на межквартильном размахе (IQR) [68]. Сначала вычисляются первый (Q1) и третий (Q3) квартили, после чего определяется IQR как разница между Q3 и Q1. Наблюдения, которые находятся ниже (Q1–1,5*IQR) или выше (Q3 + 1.5*IQR), классифицируются как выбросы. Этот метод позволяет эффективно выявлять аномалии и минимизировать их влияние на анализ данных.

Таблица 2. **Компании — лидеры по величине активов**

Table 2. **The leading companies in terms of assets**

| Компания | Активы, млрд руб. | Код ОКВЭД | Численность сотрудников |
|---|---|---|---|
| ООО «Амурский ГХК» | 464,75 | 20 | 564 |
| ООО «НМ-Тех» | 216,77 | 26 | 398 |
| ООО «РИ-Инвест» | 183,47 | 19 | 697 |

*Источник*: составлено авторами.

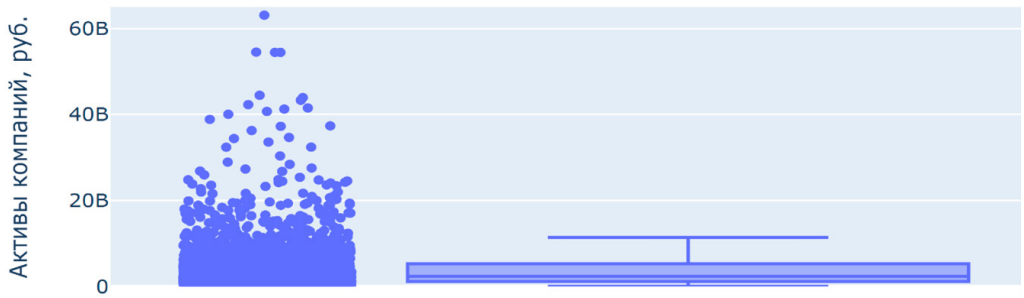


**Рис. 3.** «Ящик с усами» по активам компаний

**Fig. 3.** The "box plot" by company assets

*Источник*: расчеты авторов.

### *3.2. Методология исследования*

Методологически работа направлена на сопоставление одномерных показателей значимости признаков с модельными важностями и на количественную оценку эффекта отбора признаков на качество многоклассовой классификации. Поскольку метки получены кластеризацией в пространстве финансовых коэффициентов, основное моделирование проводится именно на коэффициентах, предварительно очищенных от выбросов и нормализованных по min-max. Данные относятся к одному календарному году, поэтому временная валидация не требуется. Для оценки обобщающей способности применяется стратифицированная k-блочная кросс-валидация с фиксированным инициализационным состоянием.

Во избежание утечки информации все операции предобработки и отбора выполняются строго на тренировочных частях фолдов. На каждом тренировочном подмножестве для числовых коэффициентов рассчитываются одномерные статистики: преобразованный в $\eta^2$ однофакторный ANOVA-F как мера различий средних между тремя классами и one-vs-rest ROC-AUC, для которого берется max(AUC, 1–AUC) с последующей взвешенной агрегацией по классам. Для всего набора признаков оценивается взаимная информация. Значения метрик нормируются по признакам методом min-max и суммируются в композитный скор, который задает ранжирование по одномерной связи и позволяет формировать топ-k кандидатов к отбору. После чего эти величины усредняются по фолдам, фиксируются ранги и их стабильность, что обеспечивает переход к сопоставлению с модельными важностями.

Далее обучается многоклассовая модель CatBoost с ранней остановкой. Качество оценивается по многоклассовой логистической функции ошибки (далее логлосс) на валидации каждого фолда. Для интерпретации на валидационных частях фолдов вычисляются ShapValues. Глобальная важность признака определяется как среднее абсолютное значение SHAP, агрегированное по наблюдениям и по классам (в макро- или взвешенной постановке), а затем усредняется по фолдам. Параллельно формируются профили важности, которые используются на этапе интерпретации в терминах финансовой логики (ликвидность, левередж, маржинальность и т. п.), после чего можно переходить к количественному сопоставлению ранжирований.

Сопоставление проводится через корреляцию рангов Спирмена между одномерным композитным ранжированием и агрегированными SHAP важностями, а также через метрики пересечения топ-k (precision@k, индекс Жаккара). Это позволяет не только оценить степень согласованности подходов, но и выявить признаки, чья значимость проявляется лишь условно — во взаимодействиях с другими факторами (высокие по SHAP при скромных одномерных метриках) или, напротив, одномерно сильные, но легко заменяемые коррелирующими переменными (высокие одномерные, низкие по SHAP). На основании этих результатов мы переходим к экспериментам по абляции.

Чтобы количественно оценить компромисс между простотой и точностью, на каждом фолде моделируются урезанные постановки: обучение на топ-k признаках, отобранных по композиту «до обучения», и обучение на топ-k по SHAP (рассчитанному на внутренней кросс валидации внутри тренировочной части фолда). Для обоих случаев логлосс на валидации сравнивается с моделью на полном наборе признаков, а различия проверяются парным критерием Уилкоксона с альтернативой «ухудшение при отборе».

Дополнительно строится кривая зависимости качества от k ($k \in \{5, 10, 15, 20, 30\}$), что позволяет определить минимальное число признаков, при котором ухудшение качества остается приемлемым. После фиксации этого порога переходят к формулировке практических рекомендаций и интерпретации ключевых коэффициентов, определяющих принадлежность к кластерам.

## 4. Результаты

### *4.1. Одномерные оценки важности*

Совокупная картина, полученная из ANOVA/$\eta^2$, Kruskal/$\varepsilon^2$, взаимной информации (MI) и одномерного AUC (табл. 3), демонстрирует, что классы компаний в первую очередь различаются по показателям платежеспособности и структуре капитала. Именно эти коэффициенты системно выходят в верхнюю часть ранжирования и дают наибольшую одномерную разделяющую способность между тремя кластерами.

Таблица 3. **Статистическая оценка топ-15 показателей**
Table 3. **Statistical evaluation of the top 15 indicators**

| № | Показатель | anova_eta2 | kruskal_eps2 | mi | ovr_auc |
|---|---|---|---|---|---|
| 1 | Коэффициент обеспеченности собственными средствами | 0,61 | 0,66 | 0,54 | 0,77 |
| 2 | Коэффициент текущей ликвидности | 0,61 | 0,69 | 0,65 | 0,76 |
| 3 | Коэффициент автономии | 0,59 | 0,62 | 0,48 | 0,79 |
| 4 | Коэффициент покрытия активов | 0,66 | 0,59 | 0,45 | 0,79 |
| 5 | Коэффициент покрытия инвестиций | 0,59 | 0,61 | 0,48 | 0,79 |
| 6 | Коэффициент обеспеченности запасов | 0,54 | 0,60 | 0,47 | 0,75 |
| 7 | Коэффициент финансовой зависимости | 0,15 | 0,38 | 0,48 | 0,77 |
| 8 | Коэффициент быстрой ликвидности | 0,48 | 0,39 | 0,26 | 0,74 |

Окончание табл. 3

| № | Показатель | anova_eta2 | kruskal_eps2 | mi | ovr_auc |
|---|---|---|---|---|---|
| 9 | Коэффициент финансового левериджа | 0,15 | 0,38 | 0,48 | 0,77 |
| 10 | Коэффициент капитализации | 0,15 | 0,38 | 0,48 | 0,77 |
| 11 | Рентабельность по чистой прибыли | 0,28 | 0,29 | 0,18 | 0,68 |
| 12 | Рентабельность продаж | 0,25 | 0,24 | 0,17 | 0,66 |
| 13 | Коэффициент покрытия процентов | 0,18 | 0,24 | 0,17 | 0,69 |
| 14 | Рентабельность активов | 0,25 | 0,25 | 0,15 | 0,66 |
| 15 | Оборачиваемость кредиторской задолженности | 0,21 | 0,23 | 0,14 | 0,67 |

*Источник*: расчеты авторов.

В лидирующей группе оказываются «коэффициент обеспеченности собственными средствами», «коэффициент текущей ликвидности», «коэффициент автономии», а также различные показатели покрытия («активов» и «инвестиций») и обеспеченности запасов. Их значения AUC на уровне порядка 0,75–0,79 при крупных эффектах $\eta^2/\varepsilon^2$ (около 0,60–0,66) указывают, что границы между «плохими», «нормальными» и «хорошими» кластерами хорошо проявляются уже в одномерных сечениях по этим признакам. Параметрические и непараметрические тесты здесь согласованы, что добавляет уверенности в устойчивости различий и снижает риск того, что результаты обусловлены особенностями распределений.

На втором плане остаются показатели прибыльности и эффективности: рентабельность по чистой прибыли и продажам, рентабельность активов, покрытие процентов и оборачиваемость кредиторской задолженности заметно ниже в ранге и демонстрируют умеренные одномерные эффекты (AUC порядка 0,66–0,69, $\eta^2$ порядка 0,18–0,28). Это косвенно подтверждает, что кластерная структура в данных отражает прежде всего платежеспособность и капитальную устойчивость, а не текущую доходность или операционную эффективность.

Характерная особенность рейтинга — «пучок» взаимосвязанных коэффициентов, описывающих один и тот же финансовый аспект с разных сторон: «финансовая зависимость», «финансовый леверидж» и «капитализация» — имеют практически совпадающие значения метрик. Такой профиль типичен для взаимно-обратных или линейно связанных показателей (например, доля собственного капитала и отношение долга к капиталу). Поэтому их целесообразно трактовать как единую факторную группу, избегая переинтерпретации каждой метрики по отдельности.

Наконец, взаимная информация вносит важную проверку на нелинейность: для лидеров по ликвидности и автономии она высокая, тогда как для показателей рентабельности заметно ниже. Это означает, что общий порядок важностей не определяется специфическими нелинейностями или экзотическими эффектами распределений. Скорее он отражает базовое различие классов по платежеспособности и структуре капитала. В дальнейшем, при переходе к модельным важностям (SHAP), естественно ожидать перераспределение вклада внутри коррелированных групп: модель оставит 1–2 представителя «капитальной» группы в топе и частично размажет вклад между дублями, а также может поднять отдельные коэффициенты (например, покрытия) за счет учета взаимодействий, которые одномерные тесты не видят.

### *4.2. Обучение модели на полном наборе показателей*

Качество многоклассовой модели вышло очень высоким: средний логлосс на пяти стратифицированных фолдах составил 0,061 ± 0,012. Это означает, что классы хорошо разделимы в пространстве финансовых коэффициентов, и дальнейшая интерпретация важностей отражает устойчивый, не шумовой сигнал (табл. 4).

Таблица 4. **Общие SHAP оценки обученной модели**
Table 4. **Statistical evaluation of the top 15 indicators**

| № | Показатель | shap | shap_std |
|---|---|---|---|
| 1 | Коэффициент обеспеченности собственными средствами | 1,014 | 0,065 |
| 2 | Коэффициент текущей ликвидности | 0,587 | 0,070 |
| 3 | Коэффициент автономии | 0,573 | 0,053 |
| 4 | Коэффициент покрытия активов | 0,354 | 0,059 |
| 5 | Коэффициент покрытия инвестиций | 0,325 | 0,040 |
| 6 | Коэффициент обеспеченности запасов | 0,280 | 0,034 |
| 7 | Коэффициент финансовой зависимости | 0,243 | 0,020 |
| 8 | Коэффициент быстрой ликвидности | 0,219 | 0,010 |
| 9 | Коэффициент финансового левериджа | 0,167 | 0,020 |
| 10 | Коэффициент капитализации | 0,148 | 0,016 |
| 11 | Рентабельность по чистой прибыли | 0,144 | 0,010 |
| 12 | Рентабельность продаж | 0,129 | 0,016 |
| 13 | Коэффициент покрытия процентов | 0,110 | 0,012 |
| 14 | Рентабельность активов | 0,108 | 0,007 |
| 15 | Оборачиваемость кредиторской задолженности | 0,103 | 0,024 |

*Источник*: расчеты авторов.

Глобальное ранжирование SHAP подтверждает центральную роль платежеспособности и ликвидности, но вносит важные акценты. Абсолютным лидером стал показатель «собственные оборотные средства»: его средняя абсолютная SHAP-важность заметно превосходит остальные признаки, причем со стабильностью по фолдам. Следом идут текущая и быстрая ликвидность, коэффициенты покрытия активов/инвестиций, обеспеченность собственными средствами, обеспеченность запасов и автономия. В середине рейтинга находятся покрытие процентов и абсолютная ликвидность, еще ниже — показатели доходности (рентабельности по чистой прибыли/продажам/активам) и операционной эффективности (оборачиваемость кредиторской задолженности, фондоотдача/фондоемкость).

«Капитальная» триада (финансовый леверидж, капитализация, финансовая зависимость) сохраняет вклад, но уже не доминирует. Модель, по сути, распределяет их общий эффект с близкими по смыслу метриками (автономия, обеспеченность) и опирается на более «операциональные» индикаторы ликвидности.

Разложение по классам уточняет картину границ. Для класса 0 («плохие») наибольший вклад в различение вносит дефицит собственных оборотных средств, за ним — текущая и быстрая ликвидность, обеспеченность собственными средствами и запасами, покрытие активов и автономия (табл. 5). Среди «второй линии» появляются абсолютная ликвидность и рентабельность по чистой прибыли. Интерпретационно это выглядит согласованно: сочетание недостатка оборотного капитала и слабой ликвидности — ключевой маркер попадания в «плохой» кластер, а прибыльность играет вспомогательную роль.

Таблица 5. **SHAP оценки обученной модели по классу 0**
Table 5. **SHAP estimates of the trained model by class 0**

| № | Показатель | shap |
|---|---|---|
| 1 | Собственные оборотные средства | 1,519 |
| 2 | Коэффициент текущей ликвидности | 0,518 |
| 3 | Коэффициент обеспеченности собственными средствами | 0,400 |
| 4 | Коэффициент обеспеченности запасов | 0,306 |
| 5 | Коэффициент быстрой ликвидности | 0,298 |
| 6 | Коэффициент покрытия активов | 0,276 |
| 7 | Коэффициент автономии | 0,176 |
| 8 | Коэффициент покрытия инвестиций | 0,173 |
| 9 | Рентабельность по чистой прибыли | 0,113 |
| 10 | Коэффициент абсолютной ликвидности | 0,102 |

*Источник*: расчеты авторов.

Для класса 1 («нормальные») профиль близок, но с некоторыми смещениями: вновь доминируют собственные оборотные средства и показатели ликвидности, тогда как из «кромочных» метрик заметнее становится покрытие процентов. Это похоже на «буферную зону» между крайними кластерами, где устойчивость обслуживания долга начинает отличать «нормальных» от соседних групп при схожем уровне базовой ликвидности (табл. 6).

Для класса 2 («хорошие») в верхних позициях оказываются быстрая и текущая ликвидность, а также покрытие активов (табл. 7). Остальной список в предоставленном фрагменте обрывается, но уже этого достаточно, чтобы увидеть закономерность: верхний кластер отделяется именно «жесткими» ликвидными метриками и способностью покрывать активы собственными источниками. Это хорошо согласуется с глобальным рейтингом и подтверждает, что границы «сверху» сформированы не столько доходностью, сколько устойчивой моментной платежеспособностью и структурой финансирования.

Важно отметить расхождение с одномерными фильтрами: показатель «собственные оборотные средства», который не был в числе абсолютных лидеров в дообучечных тестах, у модели оказывается главным драйвером разделения. Такой сдвиг типичен для случаев, когда признак проявляет себя через пороговые эффекты и взаимодействия (например, на фоне уже учтенных коэффициентов ликвидности и автономии), которые одномерные критерии не фиксируют.

Одновременно «капитально структурная» группа метрик (автономия, левередж, капитализация) в SHAP выглядит менее раздутой, чем в фильтровых ранжированиях. Их общий смысловой вклад частично «перехватывают» более точечные индикаторы платежеспособности (ликвидности, покрытия, обеспеченности).

Таблица 6. **SHAP оценки обученной модели по классу 1**
Table 6. **SHAP estimates of the trained model by class 1**

| № | Показатель | shap |
|---|---|---|
| 1 | Собственные оборотные средства | 1,332 |
| 2 | Коэффициент быстрой ликвидности | 0,577 |
| 3 | Коэффициент текущей ликвидности | 0,497 |
| 4 | Коэффициент покрытия активов | 0,283 |
| 5 | Коэффициент обеспеченности собственными средствами | 0,260 |
| 6 | Коэффициент покрытия инвестиций | 0,210 |
| 7 | Коэффициент покрытия процентов | 0,180 |
| 8 | Коэффициент автономии | 0,176 |
| 9 | Коэффициент обеспеченности запасов | 0,176 |
| 10 | Коэффициент финансового левериджа | 0,162 |

*Источник*: расчеты авторов.

Таблица 7. **SHAP оценки обученной модели по классу 2**
Table 7. **SHAP estimates of the trained model by class 2**

| № | Показатель | shap |
|---|---|---|
| 1 | Коэффициент быстрой ликвидности | 0,844 |
| 2 | Коэффициент текущей ликвидности | 0,746 |
| 3 | Коэффициент покрытия активов | 0,503 |
| 4 | Коэффициент обеспеченности запасов | 0,357 |
| 5 | Коэффициент покрытия инвестиций | 0,344 |
| 6 | Коэффициент обеспеченности собственными средствами | 0,314 |
| 7 | Коэффициент автономии | 0,307 |
| 8 | Коэффициент покрытия процентов | 0,231 |
| 9 | Коэффициент абсолютной ликвидности | 0,216 |
| 10 | Рентабельность по чистой прибыли | 0,196 |

*Источник*: расчеты авторов.

С точки зрения устойчивости интерпретации разброс важностей по фолдам невелик для ведущих признаков, что добавляет доверия к их приоритету.

Следующий логичный шаг — визуализировать зависимость SHAP для топ-признаков (beeswarm). В практическом смысле полученная картина дает простое правило для сокращенных моделей и мониторинговых панелей: удерживая один-два представителя из каждой высоко коррелированной группы и обязательно включая собственные оборотные средства вместе с метриками ликвидности и покрытия, можно сохранить основную часть разделяющей способности между «плохими», «нормальными» и «хорошими» кластерами.

### *4.3. Влияние SHAP-значений признаков*

Рис. 4 показывает, какие значения признаков повышают или понижают вероятность отнесения к классу 0. В целом картина монотонна и экономически согласована: низкие значения ключевых коэффициентов ликвидности и обеспеченности (точки синего цвета) дают крупные положительные SHAP и тем самым резко сдвигают предсказание в сторону «плохого» класса, тогда как высокие значения (красные) — в отрицательную сторону, отталкивая объект от класса 0.

Наибольший вклад несут собственные оборотные средства. При низких значениях наблюдается длинный «положительный» хвост SHAP, что означает сильный рост вероятности попасть в класс 0 именно из-за дефицита оборотного капитала. Похожий, хотя и менее экстремальный профиль у текущей и быстрой ликвидности: малые значения устойчиво повышают вероятность класса 0, большие — снижают.

К той же «защитной» группе относятся обеспеченность собственными средствами, обеспеченность запасов, покрытие активов и покрытие инвестиций. Здесь синяя часть распределений сдвинута вправо (положительные SHAP), красная — влево (отрицательные SHAP), что указывает на то, что недообеспеченность и недопокрытие активов/инвестиций — маркеры «плохого» состояния.

Автономия ведет себя аналогично: низкая доля собственного капитала увеличивает вклад в класс 0, высокая — уменьшает. Метрики доходности (рентабельность по чистой прибыли, рентабельность активов, рентабельность продаж) и абсолютная ликвидность оказывают более умеренный и локальный эффект: при низких значениях они слегка подталкивают к классу 0, при высоких — отталкивают, но размах их SHAP заметно меньше, чем у ликвидности и обеспеченности. Операционные коэффициенты (фондоотдача, оборачиваемость дебиторской задолженности) расположены в нижней части графика с небольшой амплитудой значений, то есть играют второстепенную роль для отделения «плохого» кластера.

Ширина «ройков» по верхним признакам показывает высокую вариативность их вклада — это индикатор сильных взаимодействий и пороговых зон (особенно для собственных оборотных средств и текущей ликвидности): при критически низких уровнях вероятность класса 0 растет непропорционально быстрее. В сумме визуализация подтверждает, что принадлежность к «плохому» кластеру определяется прежде всего недостатком оборотного

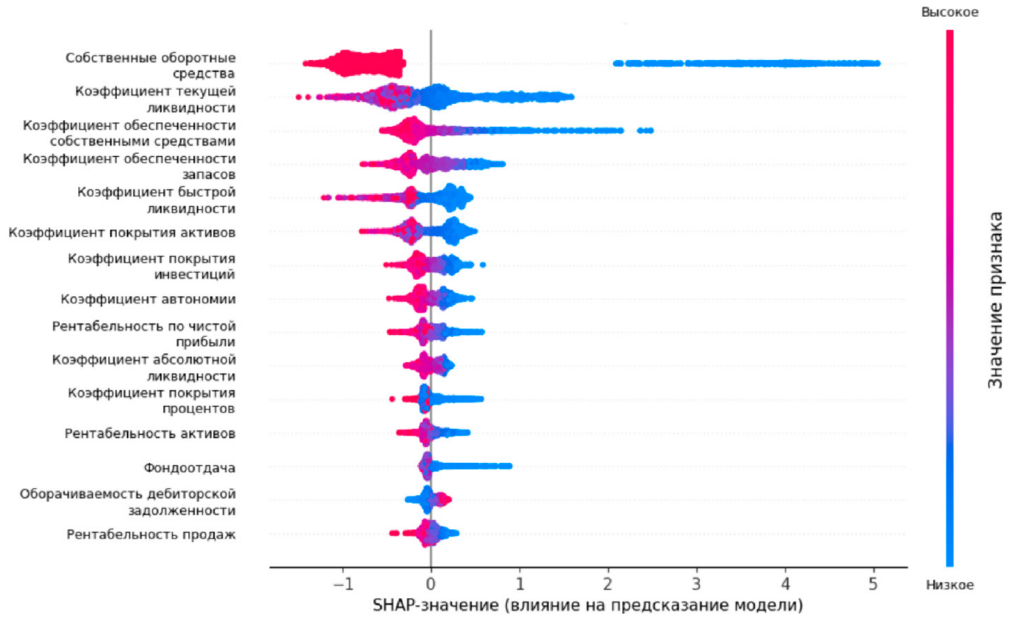


**Рис. 4.** Диаграмма «рой пчел» (beeswarm) для класса 0

**Fig. 4.** Beeswarm plot for class 0

*Источник*: расчеты авторов.

капитала и низкой ликвидностью/обеспеченностью, тогда как прибыльность и операционная эффективность играют роль уточняющих факторов.

Рис. 5 демонстрирует, какие значения признаков увеличивают или уменьшают вероятность отнесения компании к «нормальному» классу, и картина получается симметрично «промежуточной» между соседними кластерами. Самым значимым показателем оказывается поведение собственных оборотных средств. Их вклад преимущественно отрицателен для класса 1, что указывает на «колейность» нормального сегмента — как дефицит оборотного капитала, так и его избыточность уводят объект из нормального состояния к соседним классам, тогда как типичные для класса 1 уровни лежат в умеренной зоне.

На этом фоне роль ликвидности становится определяющей для удержания в классе 1: чем выше быстрая и текущая ликвидность, тем заметнее положительный сдвиг SHAP к этому классу. Наоборот, низкие значения устойчиво уменьшают вероятность «нормального» статуса и отталкивают в сторону «плохого» сегмента. В ту же сторону работают показатели покрытия (активов и инвестиций): достаточные уровни чаще поддерживают принадлежность к классу 1, тогда как недостаток покрытия систематически выбивает компанию из этой группы.

Структурные характеристики капитала — обеспеченность собственными средствами и автономия — добавляют согласованный, хотя и менее выраженный

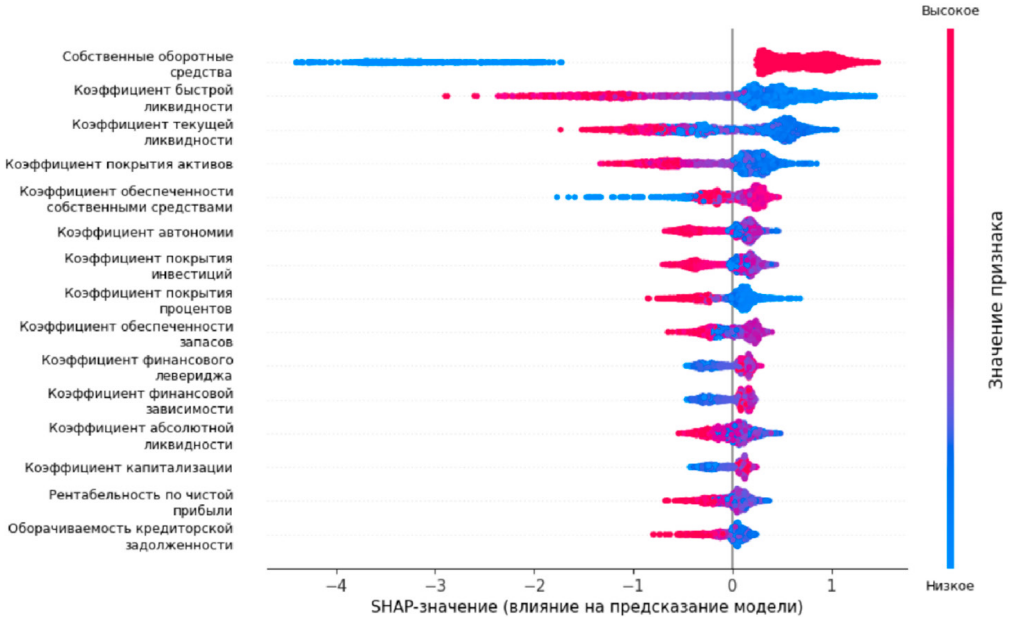


**Рис. 5.** Диаграмма «рой пчел» (beeswarm) для класса 1

**Fig. 5.** Beeswarm plot for class 1

*Источник*: расчеты авторов.

сигнал о том, что недообеспеченность и низкая доля собственного капитала снижают шансы остаться «нормальной»; приемлемые и выше среднего уровни, напротив, нейтрально или умеренно укрепляют их. Кредитный профиль проявляется ожидаемо — высокий финансовый леверидж и зависимость чаще тянут вклад в отрицательную область, тогда как достаточное покрытие процентов поддерживает принадлежность к классу 1. Иными словами, «нормальный» статус связан не с высокой долговой нагрузкой как таковой, а с ее обслуживаемостью на фоне достаточной ликвидности.

Наконец, показатели доходности и операционной эффективности (рентабельности, оборачиваемость кредиторской задолженности) вносят меньший по амплитуде вклад и играют вспомогательную роль по сравнению с блоком ликвидности, покрытия и структуры капитала. В совокупности это формирует целостный образ «нормального» класса как зоны умеренных собственных оборотных средств при достаточной моментной платежеспособности и приемлемых коэффициентах покрытия. Критически низкие уровни ликвидности и собственного капитала уводят компанию вниз, к «плохим», а экстремально высокие уровни отдельных метрик — вверх к «хорошим», что и отражается односторонней асимметрией SHAP для некоторых признаков. Такая «буферная» природа класса 1 хорошо согласуется с тем, что его границы формируются прежде всего устойчивостью платежеспособности, а не максимизацией доходности.

Диаграмма «рой пчел» для класса 2 (рис. 6), который образуют «хорошие» компании, показывает, что граница «вверх» формируется прежде всего метриками ликвидности и покрытия. Быстрая и текущая ликвидность дают самый мощный положительный вклад: по ним виден выраженный градиент цвета — высокие значения (красные) сдвигают SHAP далеко вправо, тогда как низкие (синие) — влево. Профиль практически монотонный с длинным «положительным хвостом», что говорит о пороговом характере. При достижении достаточного уровня ликвидности вероятность попасть в «хороший» класс растет резко и устойчиво.

Той же логике соответствует покрытие активов и инвестиций: достаточные значения тянут предсказание вправо, дефицит покрытия — обратно к нейтральной или отрицательной области. Обеспеченность запасов и автономия добавляют положительный, но более умеренный вклад; чем выше доля собственного капитала, тем стабильнее компания удерживается в классе 2.

Структурные показатели долговой нагрузки работают в противоположном направлении. Для финансового левериджа и финансовой зависимости заметно, что высокие значения (красные точки) чаще оказываются слева от нуля, т. е. снижают вероятность класса 2, тогда как низкие уровни долговой нагрузки не мешают положительному сдвигу. Покрытие процентов, напротив, выступает как «защитная» метрика: более высокие значения дают небольшое, но систематически положительное влияние, согласуясь с идеей обслуживаемости долга как признака «хорошего» состояния.

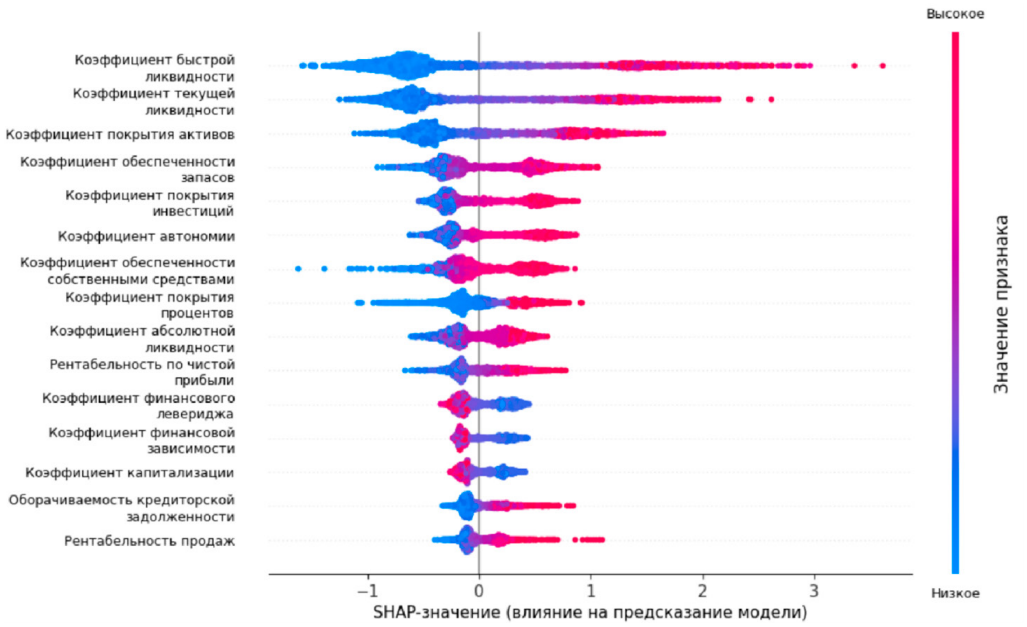


**Рис. 6.** Диаграмма «рой пчел» (beeswarm) для класса 2

**Fig. 6.** Beeswarm plot for class 2

*Источник*: расчеты авторов.

Показатели прибыльности (рентабельность по чистой прибыли, активам и продажам) усиливают принадлежность к классу 2, но их вклад уже вторичен относительно ликвидности и покрытия: красные точки смещены вправо, однако размах распределений заметно меньше, чем у лидеров. Операционные метрики, например оборачиваемость кредиторской задолженности и частично капитализационные коэффициенты, колеблются вокруг нуля с небольшой амплитудой, то есть играют вспомогательную роль.

В совокупности картина согласована экономически. «Хорошие» компании — это прежде всего те, у которых высокие быстрая и текущая ликвидность, достаточное покрытие активов/инвестиций и приемлемая автономия при невысокой долговой нагрузке и устойчивом покрытии процентов. Доходность добавляет «плюс», но не определяет класс сама по себе. По сравнению с «плохими» и «нормальными» кластерами здесь решающим становится именно качественный ликвидно-балансовый профиль, а не столько объем оборотного капитала. Модель вознаграждает устойчивую платежеспособность и структуру финансирования, что и отражается в правых, длинных хвостах SHAP по ключевым коэффициентам ликвидности и покрытия.

### *4.4. Обучение модели на топ-10 показателях по одномерной важности*

Сводка показывает, что при урезании до топ-10 признаков качество существенно падает: средний logloss возрастает с 0,0605 до 0,1439, то есть

становится примерно в 2,4 раза выше; абсолютная разница составляет +0,0834 (SD ≈ 0,015). Просадка наблюдается во всех пяти фолдах без исключений, а парный тест Уилкоксона дает p = 0,031, что указывает на статистически значимое ухудшение (табл. 8).

Интерпретационно это согласуется с предыдущими наблюдениями: заметная часть предсказательного сигнала распределена по нескольким коррелированным группам и проявляется через взаимодействия. Одномерный отбор (и в целом агрессивное сокращение до узкого топ-списка) срезает «условный» вклад признаков, который модель использует только совместно с другими. В результате даже хорошо ранжированные по фильтрам/агрегатам признаки не воспроизводят полноту разделяющей способности, которую дает полный набор.

Практически это означает следующее. Если цель — компактная модель для мониторинга, то k в районе 10 оказывается недостаточным: ожидаемо потребуется 15–20 признаков (и/или отбор по SHAP) для заметного сокращения разрыва. Если же приоритет — точность, разумнее оставлять полный набор и полагаться на регуляризацию и раннюю остановку.

### *4.5. Обучение моделей с разным количеством лучших одномерных признаков*

Кривая качества по числу признаков ведет себя ожидаемо. По мере роста k логлосс монотонно снижается и после k ≈ 20 выходит на плато, практически совпадая с моделью на полном наборе. Для ориентира: используются мультиклассовый логлосс (кросс-энтропия) и стандартная метрика качества классификации, где меньше — лучше.

На малых k заметно выигрывает отбор по SHAP: при k = 5–15 он существенно лучше фильтрового отбора, а к k ≈ 20 достигает уровня full (0,0601 против 0,0605) и дальше уже не дает ощутимого прироста.

Таблица 8. **Сравнение топ-5 показателей с полным набором**
Table 8. **Comparison of the top 5 indicators with the full set**

| Номер фолда на кросс-валидации | Логлосс на полном наборе признаков | Логлосс на топ-10 признаках | Дельта |
|---|---|---|---|
| 1 | 0,073 | 0,178 | 0,105 |
| 2 | 0,079 | 0,154 | 0,075 |
| 3 | 0,037 | 0,114 | 0,078 |
| 4 | 0,060 | 0,124 | 0,064 |
| 5 | 0,054 | 0,149 | 0,096 |

*Источник*: расчеты авторов.

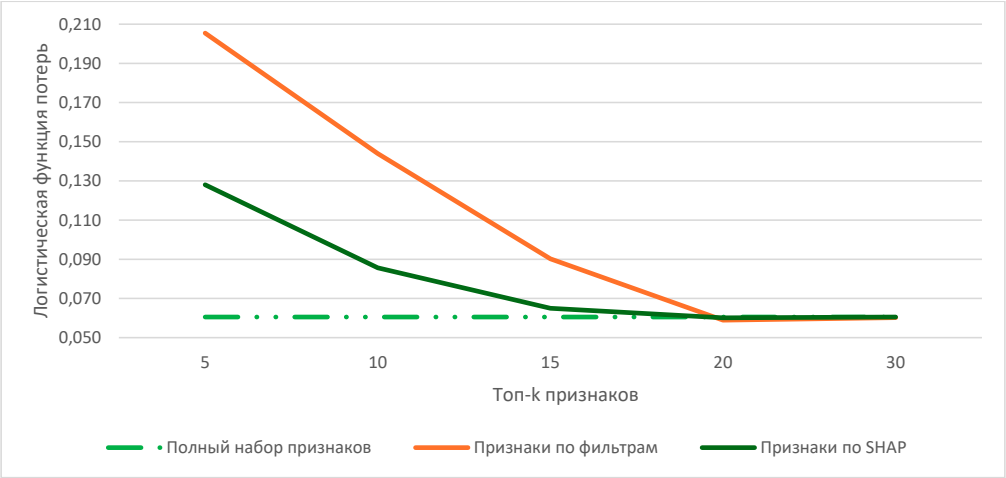


**Рис. 7.** График зависимости качества модели от выбранных признаков (меньше — лучше)

**Fig. 7.** Dependence of the model quality on the selected features (less is better)

*Источник*: расчеты авторов.

Интерпретационно это означает, что основная предсказательная масса сосредоточена примерно в 15–20 признаках. Если нужна компактная и точная модель, рационально ориентироваться на k ≈ 15–20 с SHAP-ранжированием или его пересечением с фильтрами: уже при k = 15 логлосс близок к полному, а при k ≈ 20 достигает плато. Если по организационным причинам используется только фильтровый отбор, имеет смысл подбирать k не меньше 20, чтобы нивелировать потери качества и воспользоваться «регуляризующим» эффектом сокращения. В практическом плане это также подтверждает важность работать с представителями из каждой коррелированной группы признаков: как только в списке появляются «правильные» представители ключевых групп (ликвидность, покрытие, автономия/структура капитала, собственные оборотные средства), дальнейшее расширение набора приносит все меньшую отдачу.

### *4.6. Сводный вклад показателей в предсказание модели*

Наиболее информативные показатели, отобранные из исходного набора показателей экономической безопасности предприятий и компаний представлены в табл. 9.

## 5. Обсуждение

Полученная картина согласована и с экономической логикой, и с формой построения целевой переменной. Одномерные метрики сразу выделили блок показателей платежеспособности и структуры капитала как главный источник разделимости классов, тогда как доходности и операционные коэффициенты играют второстепенную роль.

Таблица 9. **Наиболее информативные показатели экономической безопасности предприятий и компаний, отобранные на основе использования предложенного алгоритма**

Table 9. **The most informative indicators of economic security of enterprises and companies, selected based on the use of the proposed algorithm**

| Показатель / группа | Класс 0 («плохой») | Класс 1 («нормальный») | Класс 2 («хороший») |
|---|---|---|---|
| Собственные оборотные средства | Низкие значения резко повышают вероятность класса 0; очень высокая важность | «Колейность»: дефицит и избыток уменьшают шансы класса 1; умеренные уровни поддерживают; высокая важность | Не ключевой по сравнению с ликвидностью / покрытием; второстепенно |
| Ликвидность: текущая и быстрая | Низкие значения устойчиво тянут в класс 0; высокие — отталкивают; высокая важность | Определяющая роль: высокие значения повышают вероятность класса 1; низкие — снижают; высокая важность | Самые сильные положительные факторы; пороговый эффект при высоких значениях; очень высокая важность |
| Покрытие активов и инвестиций | Недостаток покрытия — маркер класса 0; высокая важность | Достаточные уровни поддерживают класс 1; дефицит — выбивает; высокая важность | Достаточные значения дают устойчиво положительный вклад; высокая важность |
| Обеспеченность собственными средствами | Низкие значения увеличивают вклад в класс 0; высокие — уменьшают; высокая важность | Недообеспеченность снижает шансы; приемлемые и выше среднего — умеренно укрепляют; средняя важность | Явный эффект не акцентирован; второстепенно относительно ликвидности/покрытия |
| Обеспеченность запасов | Недостаток — признак «плохого» состояния; средняя важность | Явно не акцентирована; второстепенно | Добавляет умеренно положительный вклад; средняя важность |
| Автономия (доля собственного капитала) | Низкая автономия повышает вероятность класса 0; высокая — снижает; средняя — высокая важность | Низкая — минус; приемлемая и выше — умеренный плюс; средняя важность | Более высокая автономия стабильно поддерживает класс 2; средняя важность |

Окончание табл. 9

| Показатель / группа | Класс 0 («плохой») | Класс 1 («нормальный») | Класс 2 («хороший») |
|---|---|---|---|
| Долговая нагрузка: финансовый леверидж / зависимость | Прямо не акцентированы; не ключевые | Высокий леверидж тянет вклад в отрицательную область; критична обслуживаемость долга; средняя важность | Высокие значения снижают вероятность класса 2; низкая долговая нагрузка не мешает положительному сдвигу; средняя важность |
| Покрытие процентов | Не ключевой | Более высокое покрытие поддерживает принадлежность к классу 1; средняя важность | Дает небольшое, но стабильное положительное влияние; второстепенно |
| Доходность (рентабельность чистой прибыли, активов, продаж) | Низкие значения слегка «подталкивают» к классу 0; высокие — отталкивают; эффект умеренный | Вспомогательная роль; малый вклад | Усиливает принадлежность, но вторична относительно ликвидности/покрытия |

*Источник*: расчеты авторов.

ISSN 2712-7435

Модельные важности по SHAP эту картину подтвердили, но важным образом уточнили акценты: абсолютным лидером стали собственные оборотные средства, затем — текущая и быстрая ликвидность, коэффициенты покрытия активов/инвестиций, обеспеченность собственными средствами и автономия. Такой сдвиг — когда признак, не доминировавший в одномерном ранге, становится главным по SHAP — типичен для случаев пороговых эффектов и взаимодействий, которые фильтры по определению не фиксируют. Одновременно «капитально структурная» триада (леверидж, финансовая зависимость, капитализация) в SHAP утратила избыточную значимость, заметную в фильтрах. Модель разделила их общий вклад между близкими по смыслу метриками и индикаторами ликвидности и покрытия, что снижает риск переинтерпретации взаимно-обратных коэффициентов как независимых факторов.

Перекладка результатов на классы показывает осмысленные, асимметричные границы. Для «плохих» компаний решающим оказывается дефицит собственных оборотных средств в сочетании с низкой текущей/быстрой ликвидностью и недостатком покрытия активов. При этих условиях SHAP-вклады крупные по модулю и преимущественно положительные, что отражает резкий сдвиг вероятности к классу 0.

Для «нормальных» компаний профиль выглядит буферным. Умеренные уровни оборотного капитала при достаточной ликвидности и обслуживаемости долга поддерживают принадлежность к классу 1, а отклонения в любую сторону — дефицит или, наоборот, чрезмерный избыток — выталкивают объект к соседним кластерам.

Для «хороших» компаний ключевыми становятся высокие быстрая и текущая ликвидность, достаточное покрытие активов/инвестиций и приемлемая автономия при невысокой долговой нагрузке. Доходности укрепляют принадлежность к классу 2, но не определяют ее самостоятельно. Такая асимметрия по знакам и «длинные хвосты» SHAP у лидирующих признаков указывают на существование пороговых зон, достижение которых резко меняет класс.

Эксперименты по сокращению признаков показывают четкий компромисс между компактностью и точностью. При жестком усечении до коротких списков SHAP-ранжирование уверенно обгоняет фильтровое: на $k = 5–15$ логлосс существенно ниже в SHAP-списках и быстро приближается к уровню полной модели. Начиная примерно с $k \approx 20$, все стратегии достигают плато и практически совпадают с качеством на полном наборе. В тоже время сокращение до $k = 10$ приводит к статистически значимому ухудшению (рост logloss на ~0,08 пункта), причем во всех фолдах без исключений. Это указывает, что заметная доля сигнала распределена по нескольким коррелированным группам признаков и проявляется через их совместное использование моделью.

Практический вывод простой: если целиться в компактность без ощутимой потери качества, разумно выбирать k в диапазоне 15–20 и опираться

на SHAP-ранг (или его пересечение с фильтрами), следя за тем, чтобы в укороченном списке были представлены ключевые смысловые группы — ликвидность (быстрая, текущая, абсолютная), собственные оборотные средства, покрытие активов/инвестиций, автономия/структура капитала и обслуживаемость долга.

С точки зрения управленческой интерпретации результаты легко транслируются в мониторинговые панели. Минимальный «обязательный» набор для раннего предупреждения о переходе в «плохой» кластер включает собственные оборотные средства, текущую и быструю ликвидность, покрытие активов (и/или инвестиций), показатель автономии и покрытие процентов; именно по ним видны наиболее выраженные и устойчивые градиенты SHAP. В дополнение имеет смысл группировать взаимозависимые коэффициенты (леверидж/капитализация/финансовая зависимость) и рассматривать их групповую важность, чтобы избежать избыточной детализации и мультиколлинеарного шума в объяснениях.

Подводя итоги, полученные результаты в целом подтверждают поставленные гипотезы. Во-первых, высокая точность многоклассовой модели (logloss 0,061 ± 0,012 на 5 стратифицированных фолдах) позволяет опираться на модельные важности для объективного ранжирования показателей, что напрямую поддерживает H1. Глобальное SHAP-ранжирование формирует четкое ядро наиболее информативных признаков: собственные оборотные средства (безусловный лидер), текущая и быстрая ликвидность, показатели покрытия активов/инвестиций, обеспеченность собственными средствами и автономия. Одномерные критерии (ANOVA/$\eta^2$, Kruskal/$\varepsilon^2$, взаимная информация, one-vs-rest ROC-AUC) также систематически поднимают в верхние позиции метрики платежеспособности и структуры капитала, что дополнительно подтверждает возможность объективной идентификации ключевых признаков для оценки экономической безопасности.

Вместе с тем сопоставление одномерных фильтров с глобальными SHAP-важностями демонстрирует лишь частичное согласие, как и предполагалось в H2: лидеры совпадают на уровне смысловых блоков (ликвидность, обеспеченность/автономия), однако порядок и акценты заметно различаются. Показательный пример — собственные оборотные средства: в одномерных тестах они не выступают абсолютным лидером, тогда как по SHAP становятся главным драйвером разделения классов, что отражает влияние взаимодействий и пороговых эффектов, невидимых одномерным критериям. Напротив, «капитально структурная» триада (леверидж, капитализация, финансовая зависимость) оказывается завышенной по одномерным фильтрам и сниженной в SHAP из-за перераспределения вклада с близкими по смыслу метриками ликвидности и обеспеченности, т. е. наблюдается именно частичное согласие.

Эту картину дополняет поведение качества при агрессивном сокращении признакового пространства и подтверждает H3: при отборе топ-10 по одномерным фильтрам logloss возрастает с 0,0605 до 0,1439 (примерно в 2,4 раза),

эффект стабилен на всех фолдах и статистически значим (Уилкоксон p = 0,031), что указывает на потерю условной (совместной) информации. В тоже время подсеты, сформированные по глобальным важностям SHAP, деградируют заметно медленнее: при малых k (5–15) они существенно превосходят одномерные отборы, а при $k \approx 20$ достигают уровня полной модели (0,0601 против 0,0605), что отражает преимущество учета совместного влияния признаков.

Наконец, влияние мультиколлинеарности согласуется с H4. В данных присутствуют тесно коррелированные группы показателей, из-за чего модель размазывает вклад между дублирующимися переменными и оставляет в топе 1–2 представителя группы; лидирующие SHAP-важности при этом демонстрируют устойчивость (малый разброс по фолдам), сохраняя информативность, но требуя осторожной интерпретации из-за распределения вклада между зависимыми переменными. В совокупности это позволяет считать все гипотезы подтвержденными.

## 6. Заключение

В соответствии с целью исследования построена интерпретируемая многоклассовая модель («плохие/нормальные/хорошие») с метками, полученными кластеризацией, реализована сопоставимая процедура ранжирования признаков (одномерные фильтры против SHAP для CatBoost), а также продемонстрирована воспроизводимая методика отбора подсетов признаков с оценкой потери качества.

В работе рассмотрена задача интерпретируемой многоклассовой классификации производственных компаний («плохие/нормальные/хорошие») в пространстве финансовых коэффициентов, рассматриваемых как показатели экономической безопасности, где метки сформированы кластеризацией. При этом сопоставлены одномерные метрики важности (ANOVA/$\eta^2$, OVR AUC, MI) с модельными важностями (SHAP) градиентного бустинга по деревьям (CatBoost), а также проведена количественная оценка цены сокращения числа признаков для качества классификации.

Результаты согласуются как между собой, так и с экономической логикой. Уже одномерные оценки показали, что классы различаются прежде всего по платежеспособности и структуре капитала: вверху рангов стабильно оказываются текущая и быстрая ликвидность, обеспеченность собственными средствами, автономия и показатели покрытия активов/инвестиций; доходности и операционные коэффициенты занимают вторичные позиции. Модельная интерпретация с помощью SHAP уточнила акценты: наибольший вклад в разделение классов вносит показатель «собственные оборотные средства», далее следуют ликвидностные метрики и коэффициенты покрытия, а «капитально структурная» триада (леверидж, капитализация, финансовая зависимость) проявляется менее избыточно, чем в фильтровых ранжированиях.

Такой сдвиг естественен: признаки, не доминирующие маргинально, могут играть ключевую роль условно — через пороговые эффекты

и взаимодействия, которые одномерные тесты не фиксируют. Профили SHAP дают осмысленные границы: «плохие» компании характеризуются дефицитом оборотного капитала и низкой ликвидностью/покрытием; «нормальные» — умеренными уровнями при достаточной обслуживаемости долга; «хорошие» — высокими ликвидностью и покрытием на фоне приемлемой автономии и невысокой долговой нагрузки. Визуализации beeswarm подтверждают наличие «критических зон», при достижении которых вероятность перехода между классами меняется резко.

Абляционные эксперименты прояснили компромисс между компактностью и точностью. При малых k отбор по SHAP уверенно превосходит фильтры; уже к $k \approx 15$ качество становится близким к полному, а при $k \approx 20$ — фактически достигает плато, не отличаясь от модели на всем наборе признаков. В то же время сокращение до $k = 10$ дает статистически значимое ухудшение качества во всех фолдах, что указывает: значимая часть предсказательного сигнала распределена по нескольким коррелированным группам и проявляется через их совместное использование моделью.

Таким образом использование рассмотренного в статье подхода позволило сформировать практические рекомендации по выбору наиболее информативных показателей экономической безопасности для различных классов предприятий и компаний. Работа вносит вклад в методологию оценки экономической безопасности: объединяет кластеризационную разметку с интерпретируемым многоклассовым машинным обучением, сопоставляет одномерные и условные важности признаков, демонстрируя, как взаимодействия и мультиколлинеарность перераспределяют вклад, и предлагает операциональный протокол измерения «цены» сокращения признакового пространства. Показано, что глобальные объяснимости (SHAP) дают более содержательную картину причинно-следственных контуров по сравнению с одномерными фильтрами, сохраняя интерпретируемость на уровне порогов и направлений эффекта.

С точки зрения практической ценности, предложенный алгоритм позволяет выстраивать объективное ранжирование показателей ЭБ для мониторинга и скоринга компаний, формировать компактные, «производственные» наборы метрик без существенной потери качества, выявлять пороговые значения ключевых коэффициентов для сценарного анализа и раннего предупреждения, учитывать мультиколлинеарность, интерпретируя группы зависимых показателей как единые блоки. Решение технологически воспроизводимо, масштабируемо на другие отрасли и может быть встроено в корпоративные и надзорные панели мониторинга.

### Список использованных источников

1. *Буланов Л. А., Калина А. В., Криворотов В. В.* Кластеризация российских производственных компаний по показателям их финансового состояния с использованием технологий машинного обучения // Journal of Applied Economic Research. 2025. Т. 24, № 2. С. 584–621. https://doi.org/10.15826/vestnik.2025.24.2.020

2. *Hauge J., Houtzager B., Hörmann A. J.* The new economic nationalism: industrial policy and national security in the United States, China, and the European Union // Geoforum. 2025. Vol. 166. 104382. https://doi.org/10.1016/j.geoforum.2025.104382

3. *Ahumada J. M.* Bringing freedom back to developmentalism: industrialisation as national independence // Cambridge Journal of Economics. 2023. Vol. 47, Issue 6. Pp. 1037–1056. https://doi.org/10.1093/cje/bead030

4. *Ghazouani T.* Geopolitical, technological, and economic influences on U.S. energy security: A quantile-wavelet analysis // Journal of Environmental Management. 2025. Vol. 390. 126393. https://doi.org/10.1016/j.jenvman.2025.126393

5. *Akusta E.* The impacts of different types of globalization on energy security risk: Can globalization be a remedy for the energy security risk of OECD countries? // Energy. 2024. Vol. 313. 133787. https://doi.org/10.1016/j.energy.2024.133787

6. *Yao L., Chang Y.* Energy security in China: A quantitative analysis and policy implications // Energy Policy. 2014. Vol. 67. Pp. 595–604. https://doi.org/10.1016/j.enpol.2013.12.047

7. *Kruyt B., van Vuuren D. P., de Vries H. J.M., Groenenberg H.* Indicators for energy security // Energy Policy. 2009. Vol. 37, Issue 6. Pp. 2166–2181. https://doi.org/10.1016/j.enpol.2009.02.006

8. *Löschel A., Moslener U., Rübbelke D. T.G.* Indicators of energy security in industrialised countries // Energy Policy. 2009. Vol. 37, Issue 6. Pp. 2166–2181. https://doi.org/10.1016/j.enpol.2009.02.006

9. *Ang B. W., Choong W. L., Ng T. S.* Energy security: Definitions, dimensions and indexes // Renewable and Sustainable Energy Reviews. 2015. Vol. 42. Pp. 1077–1093. https://doi.org/10.1016/j.rser.2014.10.064

10. *Abd El-Ghani S. S., Mansour T. G.I., Esleem S. A.* The most important economic and social indicators of the challenges facing food security for the most important crops in Egypt // Environmental and Sustainability Indicators. 2025. Vol. 25. 100808. https://doi.org/10.1016/j.indic.2025.100808

11. *Akincil O.S., Kumcu S. Y.* Towards Resilient and Sustainable Food Systems: Integrating Agricultural Production Efficiency and Food Security // Archives des Sciences: A Multidisciplinary Journal. 2024. Vol. 74, Issue 1. Pp. 51–56. https://doi.org/10.62227/as/74106

12. *Onwe J. C., Makuachukwu G. Ojide M. G., Subhan M., Forgenie D.* Food security in Nigeria amidst globalization, economic expansion, and population growth: A wavelet coherence and QARDL analysis // Journal of Agriculture and Food Research. 2024. Vol. 18. 101413. https://doi.org/10.1016/j.jafr.2024.101413

13. *Boiral O., Brotherton M.-C., Talbot D.* Anticipating the unforeseeable? ESG risk management in mining companies // Resources Policy. 2025. Vol. 106. 105628. https://doi.org/10.1016/j.resourpol.2025.105628

14. *Bertolotti A.* Effectively managing risks in an ESG portfolio // Journal of Risk Management in Financial Institutions. 2020. Vol. 13, Issue 3. Pp. 202–211. https://doi.org/10.69554/TQVN4497

15. *Razak S. E.A., Mustapha M., Shah S. M., Abu Kasim N. A.* Sustainability risk management: Are Malaysian companies ready? // Heliyon. 2024. Vol. 10, Issue 3. e24681. https://doi.org/10.1016/j.heliyon.2024.e24681

16. *Belás J., Mišanková M., Schönfeld J., Gavurová B.* Credit risk management: financial safety and sustainability aspects // Journal of Security and Sustainability Issues. 2017. Vol. 7, No. 1. Pp. 79–93. http://dx.doi.org/10.9770/jssi.2017.6.1(7)

17. *Altman E. I., Sabato G., Wilson N.* The Value of Non-Financial Information in Small and Medium-Sized Enterprise Risk Management // Journal of Credit Risk. 2010. Vol. 6, No. 2. Pp. 95–127. https://doi.org/10.21314/JCR.2010.110

18. *Behr P. Guttler A.* Credit Risk Assessment and Relationship Lending: An Empirical Analysis of German Small and Medium-Sized Enterprises // Journal of Small Business Management. 2017. Vol. 45, Issue 2. Pp. 194–213. https://doi.org/10.1111/j.1540-627X.2007.00209.x

19. *Pylypenko K. A., Babiy I. V., Volkova N. V., Feofanov L. K., Kashchena N. B.* Structuring economic security of the organization // Journal of Security and Sustainability Issues. 2019. Vol. 9, No. 1. Pp. 27–38. http://doi.org/10.9770/jssi.2019.9.1(3)

20. *Ianioglo A.* Comprehensive system of ensuring the economic security of enterprise // Agricultural and Resource Economics: International Scientific E-Journal. 2015. Vol. 1, No. 1. Pp. 69–79. https://doi.org/10.51599/are.2015.01.01.07

21. *Ianioglo A., Polajeva T.* Innovative component of economic security of enterprises: A case of the Republic of Moldova // Journal of Business Economics and Management. 2017. Vol. 18, No. 6. Pp. 1228–1242. https://doi.org/10.3846/16111699.2017.1405365

22. *Ereshko F., Karanina E.* A systematic approach to the diagnosis of economic security of a transport enterprises // Transportation Research Procedia. 2022. Vol. 63. Pp. 322–328. https://doi.org/10.1016/j.trpro.2022.06.019

23. *Ryazanova O., Timin A.* Assessment of the state, threats and risks of the economic security system of motor transport enterprises of the Kirov region // Transportation Research Procedia. 2022. Vol. 63. Pp. 1121–1130. https://doi.org/10.1016/j.trpro.2022.06.115

24. *Prievozník P., Strelcová S., Sventeková E.* Economic Security of Public Transport Provider in a Three-Dimensional Model // Transportation Research Procedia. 2021. Vol. 55. Pp. 1570–1577. https://doi.org/10.1016/j.trpro.2021.07.146

25. *Beaver W. H.* Financial Ratios As Predictors Of Failure // Journal of Accounting Research. 1966. Vol. 4. Pp. 71–111. https://doi.org/10.2307/2490171

26. *Altman E. I.* Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy // Journal of Finance. 1968. Vol. 23, No. 4. Pp. 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x

27. *Ohlson J. A.* Financial Ratios and the Probabilistic Prediction of Bankruptcy // Journal of Accounting Research. 1980. Vol. 18, No. 1. Pp. 109–131. https://doi.org/10.2307/2490395

28. *Shumway T.* Forecasting Bankruptcy More Accurately: A Simple Hazard Model // Journal of Business. 2001. Vol. 74, No. 1. Pp. 101–124. https://doi.org/10.1086/209665

29. *Lessmann S., Baesens B., Seow H.-V., Thomas L. C.* Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research // European Journal of Operational Research. 2015. Vol. 247. Pp. 124–136. https://doi.org/10.1016/j.ejor.2015.05.030

30. *Saisana M., Saltelli A., Tarantola S.* Uncertainty and Sensitivity Analysis Techniques as Tools for the Assessment of Composite Indicators // Journal of the Royal Statistical Society Series A: Statistics in Society. 2005. Vol. 168, Issue 2. Pp. 307–323. https://doi.org/10.1111/j.1467-985X.2005.00350.x

31. *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. 2017. Vol. 28, Issue 2. Pp. 131–171. https://doi.org/10.1111/jifm.12053

32. *Guyon I., Elisseeff A.* An introduction to variable and feature selection // Journal of Machine Learning Research. 2003. Vol. 3. Pp. 1157–1182. https://doi.org/10.1162/153244303322753616

33. *Saeys Y., Inza I., Larranaga P.* A review of feature selection techniques in bioinformatics // Bioinformatics. 2007. Vol. 23, Issue 19. Pp. 2507–2517. https://doi.org/10.1093/bioinformatics/btm344

34. *Li J., Cheng K., Wang S., Morstatter F., Trevino R. P., Tang J., Liu H.* Feature selection: A data perspective // ACM Computing Surveys (CSUR). 2017. Vol. 50, Issue 6. 94. https://doi.org/10.1145/3136625

35. *Kruskal W. H., Wallis W. A.* Use of ranks in one-criterion variance analysis // Journal of the American Statistical Association. 1952. Vol. 47, Issue 260. Pp. 583–621. https://doi.org/10.1080/01621459.1952.10483441

36. *Hand D. J., Till R. J.* A simple generalisation of the area under the ROC curve for multiple class classification problems // Machine Learning. 2001. Vol. 45. Pp. 171–186. https://doi.org/10.1023/A:1010920819831

37. *Fawcett T.* An introduction to ROC analysis // Pattern Recognition Letters. 2006. Vol. 27, Issue 8. Pp. 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

38. *Bademlioglu A. H., Karatas O. B., Sokmen K. F., Yuruklu E.* Thermal management and fin characteristic optimization of an electronic power supply utilizing Taguchi and ANOVA methods // Applied Thermal Engineering. 2024. Vol. 252. 123671. https://doi.org/10.1016/j.applthermaleng.2024.123671

39. *Aravind Kumar P., Sivaranjini S., JyothirGanesh P., Garlapati C.* New solubility model to correlate solubility of anticancer drugs in supercritical carbon dioxide and evaluation with Kruskal–Wallis test // Fluid Phase Equilibria. 2024. Vol. 582. 114099. https://doi.org/10.1016/j.fluid.2024.114099

40. *Kraskov A., Stögbauer H., Grassberger P.* Estimating mutual information // Physical Review E-Statistical, Nonlinear, and Soft Matter Physics. 2004. Vol. 69. 066138. https://doi.org/10.1103/PhysRevE.69.066138

41. *Ross B. C.* Mutual information between discrete and continuous data sets // PloS One. 2014. Vol. 9, Issue 2. e87357. https://doi.org/10.1371/journal.pone.0087357

42. *Zhang X., Tian R., Xu H., Liu X, C. Soares C. G.* Ship manoeuvring model identification using mutual information clustering with stepwise regression structural optimisation under small-sample constraints // Ocean Engineering. 2025. Vol. 341, Part 2. 122594. https://doi.org/10.1016/j.oceaneng.2025.122594

43. *Stuart T., Butler A., Hoffman P., Hafemeister C., Papalexi E., Mauck W. M., Hao Y., Stoeckius M., Smibert P., Satija R.* Comprehensive Integration of Single-Cell Data // Cell. 2019. Volume 177, Issue 7. Pp. 1888–1902. https://doi.org/10.1016/j.cell.2019.05.031

44. *Lundberg S. M., Erion G., Chen H., DeGrave A., Prutkin J. M., Nair B., Katz R., Himmelfarb J., Bansal N., Lee S.-I.* From local explanations to global understanding with explainable AI for trees // Nature Machine Intelligence. 2020. Vol. 2. Pp. 56–67. https://doi.org/10.1038/s42256-019-0138-9

45. *Štrumbelj E., Kononenko I.* Explaining prediction models and individual predictions with feature contributions // Knowledge and Information Systems. 2014. Vol. 41. Pp. 647–665. https://doi.org/10.1007/s10115-013-0679-x

46. *Lundberg S. M., Nair B., Vavilala M. S., Horibe M., Eisses M. J., Adams T., Liston D. E., Low D. K.-W., Newman S.-F., Kim J., Lee S.-I.* Explainable machine-learning predictions for the prevention of hypoxaemia during surgery // Nature Biomedical Engineering. 2018. Vol. 2. Pp. 749–760. https://doi.org/10.1038/s41551-018-0304-0

47. *Bussmann N., Giudici P., Marinelli D., Papenbrock J.* Explainable machine learning in credit risk management // Computational Economics. 2021. Vol. 57. Pp. 203–216. https://doi.org/10.1007/s10614-020-10042-0

48. *Aas K., Jullum M., Løland A.* Explaining individual predictions when features are dependent: More accurate approximations to Shapley values // Artificial Intelligence. 2021. Vol. 298. 103502. https://doi.org/10.1016/j.artint.2021.103502

49. *Covert I., Lundberg S., Lee S.-I.* Explaining by removing: A unified framework for model explanation // Journal of Machine Learning Research. 2021. Vol. 22, Issue 209. P. 1–90. https://doi.org/10.48550/arXiv.2011.14878

50. *Oladimeji O. O., Ayaz H., McLoughlin I., Unnikrishnan S.* Mutual information-based radiomic feature selection with SHAP explainability for breast cancer diagnosis // Results in Engineering. 2024. Vol. 24. 103071. https://doi.org/10.1016/j.rineng.2024.103071

51. *Zhou C., Wang Z., Wang X., Guo R., Zhang Z., Xiang X., Wu Y.* Deciphering the nonlinear and synergistic role of building energy variables in shaping carbon emissions: A LightGBM-SHAP framework in office buildings // Building and Environment. 2024. Vol. 266. 112035. https://doi.org/10.1016/j.buildenv.2024.112035

52. *Friedman J. H.* Greedy function approximation: A gradient boosting machine // The Annals of Statistics. 2001. Vol. 29, No. 5. Pp. 1189–1232. https://doi.org/10.1214/aos/1013203451

53. *Chen T., Guestrin C.* Xgboost: A scalable tree boosting system // KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: Association for Computing Machinery, 2016. Pp. 785–794. https://doi.org/10.1145/2939672.2939785

54. *Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T.-Y.* Lightgbm: A highly efficient gradient boosting decision tree // NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems / edited by U. von Luxburg, I. Guyon, S. Bengio, H. Wallach, R. Fergus. New York: Curran Associates Inc., 2017. Pp. 3149–3157. https://dl.acm.org/doi/10.5555/3294996.3295074

55. *Prokhorenkova L., Gusev G., Vorobev A., Dorogush A. V., Gulin A.* CatBoost: Unbiased boosting with categorical features // Advances in Neural Information Processing Systems. Vol. 31 (NeurIPS 2018) / edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett. Neural Information Processing Systems Foundation, 2018. Pp. 6638–6648. https://doi.org/10.48550/arXiv.1706.09516

56. *Zhang Y., Ren W., Lei J., Sun L., Mi Y., Chen Y.* Predicting the compressive strength of high-performance concrete via the DR-CatBoost model // Case Studies in Construction Materials. 2024. Vol. 21. e03990. https://doi.org/10.1016/j.cscm.2024.e03990

57. *Arif M., Ahmed S., Ge F., Kabir M., Khan Y. D., Yu D.-J., Thafar M.* StackACPred: Prediction of anticancer peptides by integrating optimized multiple feature descriptors with stacked ensemble approach // Chemometrics and Intelligent Laboratory Systems. 2022. Vol. 220. P. 104458. https://doi.org/10.1016/j.chemolab.2021.104458

58. *Guyon I., Weston J., Barnhill S., Vapnik V.* Gene Selection for Cancer Classification using Support Vector Machines // Machine Learning. 2002. Vol. 46. Pp. 389–422. https://doi.org/10.1023/A:1012487302797

59. *Whitney A. W.* A Direct Method of Nonparametric Measurement Selection // IEEE Transactions on Computers. 1971. Vol. 20, Issue 9. Pp. 1100–1103. https://doi.org/10.1109/T-C.1971.22341

60. *Pudil P., Novovičová J., Kittler J.* Floating search methods in feature selection // Pattern Recognition Letters. 1994. Vol. 15, Issue 11. Pp. 1119–1125. https://doi.org/10.1016/0167-8655(94)90127-9

61. *Kursa M. B., Rudnicki W. R.* Feature selection with the Boruta package // Journal of Statistical Software. 2010. Vol. 36, Issue 11. Pp. 1–13. https://doi.org/10.18637/jss.v036.i11

62. *Tibshirani R.* Regression shrinkage and selection via the lasso // Journal of the Royal Statistical Society: Series B (Methodological). 1996. Vol. 58, Issue 1. Pp. 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

63. *Zou H., Hastie T.* Addendum: Regularization and Variable Selection Via the Elastic Net // Journal of the Royal Statistical Society Series B: Statistical Methodology. 2005. Vol. 67, Issue 5. P. 768. https://doi.org/10.1111/j.1467-9868.2005.00527.x

64. *Meinshausen N., Bühlmann P.* Stability selection // Journal of the Royal Statistical Society Series B: Statistical Methodology. 2010. Vol. 72, Issue 4. Pp. 417–473. https://doi.org/10.1111/j.1467-9868.2010.00740.x

65. *Varma S., Simon R.* Bias in error estimation when using cross-validation for model selection // BMC Bioinformatics. 2006. Vol. 7. 91. https://doi.org/10.1186/1471-2105-7-91

66. *Cawley G. C., Talbot N. L.* On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation // The Journal of Machine Learning Research. 2010. Vol. 11. Pp. 2079–2107. https://doi.org/10.5555/1756006.1859921

67. *Wilcoxon F.* Individual Comparisons by Ranking Methods // Biometrics Bulletin. 1945. Vol. 1, No. 6. Pp. 80–83. https://doi.org/10.2307/3001968

68. Understanding Robust and Exploratory Data Analysis / edited by D. C. Hoaglin, F. Mosteller, J. W. Tukey. New York: Wiley, 1983. 447 p. URL: https://archive.org/details/understandingrob0000unse/page/n1/mode/1up

## ИНФОРМАЦИЯ ОБ АВТОРАХ

**Буланов Лев Алексеевич**

Аспирант кафедры экономической безопасности производственных комплексов Уральского федерального университета имени первого Президента России Б. Н. Ельцина, г. Екатеринбург, Россия (620002, г. Екатеринбург, ул. Мира, 19); ORCID https://orcid.org/0009-0001-0242-0127 e-mail: levbulanov2013@yandex.ru

**Калина Алексей Владимирович**

Кандидат технических наук, доцент кафедры экономической безопасности производственных комплексов Уральского федерального университета имени первого Президента России Б. Н. Ельцина, г. Екатеринбург, Россия (620002, г. Екатеринбург, ул. Мира, 19), старший научный сотрудник Центра экономической безопасности Института экономики Уральского отделения РАН, г. Екатеринбург, Россия (620014, г. Екатеринбург, ул. Московская, 29); ORCID https://orcid.org/0000-0003-0376-2505 e-mail: alexkalina74@yandex.ru

**Криворотов Вадим Васильевич**

Доктор экономических наук, профессор, заведующий кафедрой экономической безопасности производственных комплексов Уральского федерального университета имени первого Президента России Б. Н. Ельцина, г. Екатеринбург, Россия (620002, г. Екатеринбург, ул. Мира, 19); ORCID https://orcid.org/0000-0002-7066-0325 e-mail: v_krivorotov@mail.ru

## БЛАГОДАРНОСТИ

Статья выполнена в соответствии с планом научно-исследовательских работ Института экономики Уральского отделения Российской академии наук.

## ДЛЯ ЦИТИРОВАНИЯ

Буланов Л. А., Калина А. В., Криворотов В. В. Отбор информативных показателей для оценки экономической безопасности российских компаний // Journal of Applied Economic Research. 2025. Т. 24, № 4. С. 1371–1415. https://doi.org/10.15826/vestnik.2025.24.4.045

## ИНФОРМАЦИЯ О СТАТЬЕ

Дата поступления 2 сентября 2025 г.; дата поступления после рецензирования 9 октября 2025 г.; дата принятия к печати 23 октября 2025 г.

# Selection of Informative Indicators for Assessing the Economic Security of Russian Companies

***Lev A. Bulanov*** [1], ***Alexei V. Kalina*** [1, 2] ✉, ***Vadim V. Krivorotov*** [1]

[1] *Ural Federal University named after the First President of Russia B. N. Yeltsin, Yekaterinburg, Russia*

[2] *Institute of Economics, The Ural Branch of Russian Academy of Sciences, Yekaterinburg, Russia*

✉ *alexkalina74@yandex.ru*

***Abstract.*** Ensuring the economic security of enterprises and companies is one of the top priorities under the current economic and political conditions. Central to solving this problem is a comprehensive and objective assessment of the economic security of companies, which is usually based on the method of indicative analysis. This, in turn, necessitates the development and formation of indicators of economic security, which would most comprehensively and qualitatively reflect the state of companies both in the current and future periods. The purpose of this article is the development of an algorithm and its model implementation for selecting the most significant indicators of economic security of enterprises based on the use of machine learning technologies. To achieve this goal, the paper explores the interpretable multiclass classification of large manufacturing companies ("bad/normal/good") according to their financial coefficients with the formation of target labels by K-Means++ clustering on a real dataset for 2023 (2,249 observations). One-dimensional filters (ANOVA/n2, one-vs-rest ROC-AUC, Mutual Information) and model SHAP values for CatBoost are compared; the comparison is carried out through correlation of ranks, intersection of top-k and ablation with a logloss estimate on a stratified k-fold CV with strict separation of train/val. Additionally, the compromise "compactness vs accuracy" at k is evaluated. {5, 10, 15, 20, 30} and the statistical significance of the differences is checked (the Wilcoxon criterion). The proposed approach and its implementation for Russian companies have revealed that their own working capital is the basis for the formation and evaluation of the most significant and informative indicators of the companies' economic security. At the same time, class profiles show asymmetric, economically consistent threshold: for "bad" companies, the decisive factor is a shortage of working capital and low liquidity; for "good" companies, high fast/current liquidity and coverage; "normal" companies are formed as a buffer with moderate values and serviced debt. The primary practical result of the research is a set of actionable recommendations as to the minimum set of indicators and a strategy for selecting features for monitoring the economic security and financial condition of companies and applied financial analytics.

***Key words:*** economic security indicators; machine learning; cross-validation; feature selection; ANOVA; ROC-AUC; Mutual Information; CatBoost; SHAP.

**JEL** D22, G30, C45

## References

1. Bulanov, L.A., Kalina, A.V., Krivorotov, V.V. (2025). Clustering of Russian Manu-facturing Companies by Indicators of Their Financial Condition Using Machine Learning Technologies. *Journal of Applied Economic Research,* Vol. 24, No. 2, 584–621. (In Russ.). https://doi.org/10.15826/vestnik.2025.24.2.020

2. Hauge, J., Houtzager, B., Hörmann, A.J. (2025). The new economic nationalism: industrial policy and national security in the United States, China, and the European Union. *Geoforum,* Vol. 166, 104382. https://doi.org/10.1016/j.geoforum.2025.104382

3. Ahumada, J.M. (2023). Bringing freedom back to developmentalism: industrialisation as national independence. *Cambridge Journal of Economics,* Vol. 47, Issue 6, 1037–1056. https://doi.org/10.1093/cje/bead030

4. Ghazouani, T. (2025). Geopolitical, technological, and economic influences on U.S. energy security: A quantile-wavelet analysis. *Journal of Environmental Management,* Vol. 390, 126393. https://doi.org/10.1016/j.jenvman.2025.126393

5. Akusta, E. (2024). The impacts of different types of globalization on energy security risk: Can globalization be a remedy for the energy security risk of OECD countries? *Energy,* Vol. 313, 133787. https://doi.org/10.1016/j.energy.2024.133787

6. Yao, L., Chang, Y. (2014). Energy security in China: A quantitative analysis and policy implications. *Energy Policy,* Vol. 67, 595–604. https://doi.org/10.1016/j.enpol.2013.12.047

7. Kruyt, B., van Vuuren, D.P., de Vries, H.J.M., Groenenberg, H. (2009). Indicators for energy security. *Energy Policy,* Vol. 37, Issue 6, 2166–2181. https://doi.org/10.1016/j.enpol.2009.02.006

8. Löschel, A., Moslener, U., Rübbelke, D.T.G. (2009). Indicators of energy security in industrialised countries. *Energy Policy,* Vol. 37, Issue 6, 2166–2181. https://doi.org/10.1016/j.enpol.2009.02.006

9. Ang, B.W., Choong, W.L., Ng, T.S. (2015). Energy security: Definitions, dimensions and indexes. *Renewable and Sustainable Energy Reviews,* Vol. 42, 1077–1093. https://doi.org/10.1016/j.rser.2014.10.064

10. Abd El-Ghani, S.S., Mansour, T.G.I., Esleem, S.A. (2025). The most important economic and social indicators of the challenges facing food security for the most important crops in Egypt. *Environmental and Sustainability Indicators,* Vol. 25, 100808. https://doi.org/10.1016/j.indic.2025.100808

11. Akinci1, O.S., Kumcu, S.Y. (2024). Towards Resilient and Sustainable Food Systems: Integrating Agricultural Production Efficiency and Food Security. *Archives des Sciences: A Multidisciplinary Journal,* Vol. 74, Issue 1, 51–56. https://doi.org/10.62227/as/74106

12. Onwe, J.C., Makuachukwu, G. Ojide, M.G., Subhan, M., Forgenie, D. (2024). Food security in Nigeria amidst globalization, economic expansion, and population growth: A wavelet coherence and QARDL analysis. *Journal of Agriculture and Food Research,* Vol. 18, 101413. https://doi.org/10.1016/j.jafr.2024.101413

13. Boiral, O., Brotherton, M.-C., Talbot, D. (2025). Anticipating the unforeseeable? ESG risk management in mining companies. *Resources Policy,* Vol. 106, 105628. https://doi.org/10.1016/j.resourpol.2025.105628

14. Bertolotti, A. (2020). Effectively managing risks in an ESG portfolio. *Journal of Risk Management in Financial Institutions,* Vol. 13, Issue 3, 202–211. https://doi.org/10.69554/TQVN4497

15. Razak, S.E.A., Mustapha, M., Shah, S.M., Abu Kasim, N.A. (2024). Sustainability risk management: Are Malaysian companies ready? *Heliyon,* Vol. 10, Issue 3, e24681. https://doi.org/10.1016/j.heliyon.2024.e24681

16. Belás, J., Mišanková, M., Schönfeld, J., Gavurová, B. (2017). Credit risk management: financial safety and sustainability aspects. *Journal of Security and Sustainability Issues,* Vol. 7, No. 1, 79–93. http://dx.doi.org/10.9770/jssi.2017.6.1(7)

17. Altman, E.I., Sabato, G., Wilson, N. (2010). The Value of Non-Financial Information in Small and Medium-Sized Enterprise Risk Management. *Journal of Credit Risk,* Vol. 6, No. 2, 95–127. https://doi.org/10.21314/JCR.2010.110

18. Behr, P. Guttler, A. (2017). Credit Risk Assessment and Relationship Lending: An Empirical Analysis of German Small and Medium-Sized Enterprises. *Journal of Small Business Management,* Vol. 45, Issue 2, 194–213. https://doi.org/10.1111/j.1540-627X.2007.00209.x

19. Pylypenko, K.A., Babiy, I.V., Volkova, N.V., Feofanov, L.K., Kashchena, N.B. (2019). Structuring economic security of the organization. *Journal of Security and Sustainability Issues,* Vol. 9, No. 1, 27–38. http://doi.org/10.9770/jssi.2019.9.1(3)

20. Ianioglo, A. (2015). Comprehensive system of ensuring the economic security of enterprise. *Agricultural and Resource Economics: International Scientific E-Journal,* Vol. 1, No. 1, 69–79. https://doi.org/10.51599/are.2015.01.01.07

21. Ianioglo, A., Polajeva, T. (2017). Innovative component of economic security of enterprises: A case of the Republic of Moldova. *Journal of Business Economics and Management,* Vol. 18, No. 6, 1228–1242. https://doi.org/10.3846/16111699.2017.1405365

22. Ereshko, F., Karanina, E. (2022). A systematic approach to the diagnosis of economic security of a transport enterprises. *Transportation Research Procedia,* Vol. 63, 322–328. https://doi.org/10.1016/j.trpro.2022.06.019

23. Ryazanova, O., Timin, A. (2022). Assessment of the state, threats and risks of the economic security system of motor transport enterprises of the Kirov region. *Transportation Research Procedia,* Vol. 63, 1121–1130. https://doi.org/10.1016/j.trpro.2022.06.115

24. Prievozník, P., Strelcová, S., Sventeková, E. (2021). Economic Security of Public Transport Provider in a Three-Dimensional Model. *Transportation Research Procedia,* Vol. 55, 1570–1577. https://doi.org/10.1016/j.trpro.2021.07.146

25. Beaver, W.H. (1966). Financial Ratios as Predictors of Failure. *Journal of Accounting Research,* Vol. 4, 71–111. https://doi.org/10.2307/2490171

26. Altman, E.I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance,* Vol. 23, No. 4, 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x

27. Ohlson, J.A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research,* Vol. 18, No. 1, 109–131. https://doi.org/10.2307/2490395

28. Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *Journal of Business,* Vol. 74, No. 1, 101–124. https://doi.org/10.1086/209665

29. Lessmann, S., Baesens, B., Seow, H.-V., Thomas, L.C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research,* Vol. 247, 124–136. https://doi.org/10.1016/j.ejor.2015.05.030

30. Saisana, M., Saltelli, A., Tarantola, S. (2005). Uncertainty and Sensitivity Analysis Techniques as Tools for the Assessment of Composite Indicators. *Journal of the Royal Statistical Society Series A: Statistics in Society,* Vol. 168, Issue 2, 307–323. https://doi.org/10.1111/j.1467-985X.2005.00350.x

31. Altman, E.I., Iwanicz-Drozdowska, M., Laitinen, E.K., Suvas, A. (2017). Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model. *Journal of International Financial Management & Accounting,* Vol. 28, Issue 2, 131–171. https://doi.org/10.1111/jifm.12053

32. Guyon, I., Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research,* Vol. 3, 1157–1182. https://doi.org/10.1162/153244303322753616

33. Saeys, Y., Inza, I., Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics,* Vol. 23, Issue 19, 2507–2517. https://doi.org/10.1093/bioinformatics/btm344

34. Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R.P., Tang, J., Liu, H. (2017). Feature selection: A data perspective. *ACM Computing Surveys (CSUR),* Vol. 50, Issue 6, 94. https://doi.org/10.1145/3136625

35. Kruskal, W.H., Wallis, W.A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association,* Vol. 47, Issue 260, 583–621. https://doi.org/10.1080/01621459.1952.10483441

36. Hand, D.J., Till, R.J. (2001). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning,* Vol. 45, 171–186. https://doi.org/10.1023/A:1010920819831

37. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters,* Vol. 27, Issue 8, 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

38. Bademlioglu, A.H., Karatas, O.B., Sokmen, K.F., Yuruklu, E. (2024). Thermal management and fin characteristic optimization of an electronic power supply utilizing Taguchi and ANOVA methods. *Applied Thermal Engineering,* Vol. 252, 123671. https://doi.org/10.1016/j.applthermaleng.2024.123671

39. AravindKumar, P., Sivaranjini, S., JyothirGanesh, P., Garlapati, C. (2024). New solubility model to correlate solubility of anticancer drugs in supercritical carbon dioxide and evaluation with Kruskal–Wallis test. Fluid Phase Equilibria, Vol. 582, 114099. https://doi.org/10.1016/j.fluid.2024.114099

40. Kraskov, A., Stögbauer, H., Grassberger, P. (2004). Estimating mutual information. *Physical Review E-Statistical, Nonlinear, and Soft Matter Physics,* Vol. 69, 066138. https://doi.org/10.1103/PhysRevE.69.066138

41. Ross, B.C. (2014). Mutual information between discrete and continuous data sets. *PloS One,* Vol. 9, Issue 2, e87357. https://doi.org/10.1371/journal.pone.0087357

42. Zhang, X., Tian, R., Xu, H., Liu, X, C. Soares, C.G. (2025). Ship manoeuvring model identification using mutual information clustering with stepwise regression structural optimisation under small-sample constraints. *Ocean Engineering,* Vol. 341, Part 2, 122594. https://doi.org/10.1016/j.oceaneng.2025.122594

43. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell,* Volume 177, Issue 7, 1888–1902. https://doi.org/10.1016/j.cell.2019.05.031

44. Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence,* Vol. 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9

45. Štrumbelj, E., Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems,* Vol. 41, 647–665. https://doi.org/10.1007/s10115-013-0679-x

46. Lundberg, S.M., Nair, B., Vavilala, M.S., Horibe, M., Eisses, M.J., Adams, T., Liston, D.E., Low, D.K.-W., Newman, S.-F., Kim, J., Lee, S.-I. (2018). Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering,* Vol. 2, 749–760. https://doi.org/10.1038/s41551-018-0304-0

47. Bussmann, N., Giudici, P., Marinelli, D., Papenbrock, J. (2021). Explainable machine learning in credit risk management. *Computational Economics,* Vol. 57, 203–216. https://doi.org/10.1007/s10614-020-10042-0

48. Aas, K., Jullum, M., Løland, A. (2021). Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *Artificial Intelligence,* Vol. 298, 103502. https://doi.org/10.1016/j.artint.2021.103502

49. Covert, I., Lundberg, S., Lee, S.-I. (2021). Explaining by removing: A unified framework for model explanation. Journal of Machine Learning Research, Vol. 22, Issue 209, 1–90. https://doi.org/10.48550/arXiv.2011.14878

50. Oladimeji, O.O., Ayaz, H., McLoughlin, I., Unnikrishnan, S. (2024). Mutual information-based radiomic feature selection with SHAP explainability for breast cancer diagnosis. *Results in Engineering,* Vol. 24, 103071. https://doi.org/10.1016/j.rineng.2024.103071

51. Zhou, C., Wang, Z., Wang, X., Guo, R., Zhang, Z., Xiang, X., Wu, Y. (2024). Deciphering the nonlinear and synergistic role of building energy variables in shaping carbon emissions: A LightGBM-SHAP framework in office buildings. *Building and Environment,* Vol. 266, 112035. https://doi.org/10.1016/j.buildenv.2024.112035

52. Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics,* Vol. 29, No. 5, 1189–1232. https://doi.org/10.1214/aos/1013203451

53. Chen, T., Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery*

*and Data Mining.* New York, Association for Computing Machinery, 785–794. https://doi.org/10.1145/2939672.2939785

54. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems.* Edited by U. von Luxburg, I. Guyon, S. Bengio, H. Wallach, R. Fergus. New York, Curran Associates Inc., 3149–3157. https://dl.acm.org/doi/10.5555/3294996.3295074

55. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems,* Vol. 31 (NeurIPS 2018). Edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett. Neural Information Processing Systems Foundation, 6638–6648. https://doi.org/10.48550/arXiv.1706.09516

56. Zhang, Y., Ren, W., Lei, J., Sun, L., Mi, Y., Chen, Y. (2024). Predicting the compressive strength of high-performance concrete via the DR-CatBoost model. *Case Studies in Construction Materials,* Vol. 21, e03990. https://doi.org/10.1016/j.cscm.2024.e03990

57. Arif, M., Ahmed, S., Ge, F., Kabir, M., Khan, Y.D., Yu, D.-J., Thafar, M. (2022). StackACPred: Prediction of anticancer peptides by integrating optimized multiple feature descriptors with stacked ensemble approach. *Chemometrics and Intelligent Laboratory Systems,* Vol. 220, 104458. https://doi.org/10.1016/j.chemolab.2021.104458

58. Guyon, I., Weston, J., Barnhill, S., Vapnik, V. (2002). Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning,* Vol. 46, 389–422. https://doi.org/10.1023/A:1012487302797

59. Whitney, A.W. (1971). A Direct Method of Nonparametric Measurement Selection. *IEEE Transactions on Computers,* Vol. 20, Issue 9, 1100–1103. https://doi.org/10.1109/T-C.1971.22341

60. Pudil, P., Novovičová, J., Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognition Letters,* Vol. 15, Issue 11, 1119–1125. https://doi.org/10.1016/0167-8655(94)90127-9

61. Kursa, M.B., Rudnicki, W.R. (2010). Feature selection with the Boruta package. *Journal of Statistical Software,* Vol. 36, Issue 11, 1–13. https://doi.org/10.18637/jss.v036.i11

62. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological),* Vol. 58, Issue 1, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

63. Zou, H., Hastie, T. (2005). Addendum: Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology,* Vol. 67, Issue 5, 768. https://doi.org/10.1111/j.1467-9868.2005.00527.x

64. Meinshausen, N., Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society Series B: Statistical Methodology,* Vol. 72, Issue 4, 417–473. https://doi.org/10.1111/j.1467-9868.2010.00740.x

65. Varma, S., Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics,* Vol. 7, 91. https://doi.org/10.1186/1471-2105-7-91

66. Cawley, G.C., Talbot, N.L. (2010). On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *The Journal of Machine Learning Research,* Vol. 11, 2079–2107. https://doi.org/10.5555/1756006.1859921

67. Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin,* Vol. 1, No. 6, 80–83. https://doi.org/10.2307/3001968

68. Wiley (1983). *Understanding Robust and Exploratory Data Analysis.* Edited by D. C. Hoaglin, F. Mosteller, J. W. Tukey. New York, Wiley, 447 p. Available at: https://archive.org/details/understandingrob0000unse/page/n1/mode/1up

## INFORMATION ABOUT AUTHORS

**Lev Alexeevich Bulanov**

Post-Graduate Student, Department of Economic Safety of Industrial Complexes, Ural Federal University named after the first President of Russia B. N. Yeltsin, Yekaterinburg, Russia (620002, Yekaterinburg, Mira street, 19); ORCID https://orcid.org/0009-0001-0242-0127 e-mail: levbulanov2013@yandex.ru

**Alexei Vladimirovich Kalina**

Candidate of Technical Sciences, Associate Professor, Department of Economic Safety of Industrial Complexes, Ural Federal University named after the first President of Russia B. N. Yeltsin, Yekaterinburg, Russia (620002, Yekaterinburg, Mira street, 19), Senior Researcher, The Center of Economic Security, Institute of Economics, The Ural Branch of Russian Academy of Sciences, Yekaterinburg, Russia (620014, Yekaterinburg, Moskovskaya street, 29); ORCID https://orcid.org/0000-0003-0376-2505 e-mail: alexkalina74@yandex.ru

**Vadim Vasilyevich Krivorotov**

Doctor of Economics, Professor, Head of Department of Economic Safety of Indus-trial Complexes, Ural Federal University named after the first President of Russia B. N. Yeltsin, Yekaterinburg, Russia (620002, Yekaterinburg, Mira street, 19); ORCID https://orcid.org/0000-0002-7066-0325 e-mail: v_krivorotov@mail.ru

## ACKNOWLEDGMENTS

The article was carried out in accordance with the research plan of the Institute of Economics of the Ural Branch of Russian Academy of Sciences.

## FOR CITATION

Bulanov, L.A., Kalina, A.V., Krivorotov, V.V. (2025). Selection of Informative Indicators for Assessing the Economic Security of Russian Companies. *Journal of Applied Economic Research,* Vol. 24, No. 4, 1371–1415. https://doi.org/10.15826/vestnik.2025.24.4.045

## ARTICLE INFO

Received September 2, 2025; Revised October 9, 2025; Accepted October 23, 2025.