



Машинное обучение в предынвестиционной оценке инновационных компаний: объяснимый многомерный скоринг инвестиционной привлекательности и риска несостоятельности

Балтян М.А., студент кафедры «Инновационное предпринимательство» (ИБМ7),
Московский государственный технический университет имени Н.Э. Баумана,
Москва, Россия

Аннотация. Разработана и эмпирически валидирована методика многомерной объяснимой предынвестиционной оценки инновационных компаний на основе машинного обучения и открытых данных. На массиве более 5 тыс. отечественных компаний обучены модели градиентного бустинга по шести измерениям инвестиционной привлекательности с интеграцией 58 финансовых коэффициентов и моделей прогнозирования несостоятельности Альтмана и Таффлера. Применена двухуровневая валидация: воспроизведение экспертной методики ($R^2 = 0,968$) и независимый форвардный прогноз выручки ($R^2 = 0,77$); индикаторы Альтмана и Таффлера статистически разделяют действующие и проблемные компании. Показано, что автоматизация снижает издержки первичного скрининга на порядки при окупаемости разовых вложений на единицах проектов.

Ключевые слова: предынвестиционная оценка; скоринг инновационных компаний; машинное обучение; объяснимый искусственный интеллект; прогнозирование несостоятельности; инвестиционная привлекательность; открытые данные.

Machine learning in pre-investment assessment of innovative companies: explainable multidimensional scoring of investment attractiveness and insolvency risk

Baltyan M.A., student of the Department of Innovative Entrepreneurship (IBM7),
Bauman Moscow State Technical University, Moscow, Russia

Annotation. A methodology for multidimensional, explainable pre-investment assessment of innovative companies based on machine learning and open data is developed and empirically validated. Using a dataset of more than 5,000 domestic companies, gradient boosting models are trained across six dimensions of investment attractiveness, integrating 58 financial ratios and the Altman and Taffler bankruptcy-prediction models. A two-level validation is applied: reproduction of an expert methodology ($R^2 = 0.968$) and an independent forward revenue forecast ($R^2 = 0.77$); the Altman and Taffler indicators statistically separate operating and distressed companies. It is shown that automation reduces primary-screening costs by orders of magnitude, with payback achieved within a few projects.

Key words: pre-investment assessment; scoring of innovative companies; machine learning; explainable artificial intelligence; insolvency prediction; investment attractiveness; open data.

Введение. Российский рынок венчурного финансирования переживает период структурной трансформации. По данным исследования медиа «Русский Венчур» и АО «Малина венчурс» (Malina VC) при участии фонда Brio Capital, за двенадцать месяцев (с 1 декабря 2024 г. по 30 ноября 2025 г.) совокупный объём публичных венчурных сделок с участием отечественных стартапов сократился на 10,18 % — с 8,0 до 7,2 млрд руб., а число сделок снизилось со 113 до 78 [7]. Одновременно изменяется структура спроса на капитал: по оценке Московского инновационного кластера, в первом полугодии 2025 г. объём венчурных инвестиций вырос на 81 % год к году и достиг 83 млн долл. при 62 сделках, при

этом средний чек увеличился на 90 % — до 1,4 млн долл., а медианный — на 156 %, до 0,3 млн долл. [11]. Эти тенденции означают «укрупнение» сделок и переход инвесторов к более избирательной, ориентированной на зрелые и предсказуемые проекты модели поведения. В таких условиях возрастает цена ошибки отбора и, следовательно, ценность качественного предынвестиционного анализа.

Ключевая экономическая проблема предынвестиционной стадии — высокая информационная асимметрия между основателем и инвестором и связанные с ней транзакционные издержки отбора. Чтобы выявить в пуле из сотен и тысяч компаний несколько действительно перспективных, инвестор вынужден проводить дорогостоящую комплексную проверку (due diligence). По данным открытых коммерческих предложений российских аудиторско-консалтинговых компаний, стоимость профессионального due diligence одного проекта составляет от 50–120 тыс. руб. за экспресс-проверку до 250–500 тыс. руб. и более за комплексный анализ (вплоть до нескольких миллионов рублей для крупных сделок), а сроки варьируются от одной–трёх недель до полутора–двух месяцев и определяются трудозатратами в 80–300 человеко-часов [1]. Подобный подход не масштабируется: инвестор, рассматривающий десятки проектов ежемесячно, объективно не способен обеспечить каждому из них качественный анализ, в результате чего значительная часть потенциально успешных проектов отсеивается на стадии первичного скрининга без содержательной оценки.

Существующие на рынке инструменты лишь частично закрывают эту потребность. Информационно-справочные системы (СПАРК-Интерфакс, Контур.Фокус) предоставляют юридические и финансовые сведения по подписке (от 32–99 тыс. руб. в год до порядка 25 тыс. руб. в месяц) [8], но не оценивают инвестиционный потенциал; каталоги стартапов не содержат аналитики, а венчурные платформы ориентированы на ручной скаутинг. Зарубежные платформы венчурной аналитики (Harmonic.ai, PitchBook, Tracxn, Specter) реализуют скаутинг на основе машинного обучения, однако распространяются по корпоративным

тарифам в десятки тысяч долларов в год и слабо охватывают российскую юрисдикцию и национальную систему бухгалтерской отчётности [16]. Таким образом, ниша доступного инструмента, объединяющего машинное обучение, объяснимость результатов и опору на отечественные открытые данные, остаётся незанятой.

Объект исследования — процесс предынвестиционной оценки (скрининга) инновационных компаний на венчурном рынке. **Предмет исследования** — методы и инструменты автоматизированной оценки инвестиционной привлекательности и рисков инновационных компаний на основе машинного обучения и открытых данных. **Цель исследования** — разработать и эмпирически валидировать методику многомерной объяснимой предынвестиционной оценки инновационных компаний и оценить её экономический эффект. Для достижения цели поставлены задачи: (1) проанализировать методы предынвестиционной оценки и сегмент аналитических инструментов венчурного рынка, выявив незакрытую нишу; (2) разработать методику многомерного объяснимого скоринга с интеграцией финансово-аналитического блока; (3) эмпирически валидировать методику двумя независимыми способами и сравнить с базовыми алгоритмами; (4) оценить экономический эффект автоматизации скрининга.

Под инновационной компанией в настоящей работе понимается технологическая компания, находящаяся на стадии инновационного развития, обладающая измеримыми уровнями готовности технологии (в частности, по шкале TRL и производным от неё) и являющаяся резидентом институтов инновационного развития или объектом их экспертизы, вне зависимости от возраста; частным случаем такой компании является технологический стартап. Научная новизна работы состоит, во-первых, в авторской методике многомерной (шесть измерений) объяснимой предынвестиционной оценки, объединяющей шкалы готовности технологии с финансово-аналитическим блоком (коэффициенты ликвидности, устойчивости и рентабельности, а также адаптированные модели Альтмана и

Таффлера); во-вторых, в двухуровневой схеме валидации, отделяющей внутреннюю согласованность скоринга от его реальной предсказательной способности за счёт независимого форвардного финансового прогноза; в-третьих, в параметрической модели оценки экономического эффекта автоматизации предынвестиционного скрининга.

Литературный обзор. Теоретико-методологическую основу исследования формируют три направления научной литературы.

Оценка инвестиционной привлекательности технологических компаний. Классические подходы — дисконтирование денежных потоков и метод рыночных мультипликаторов — на ранних стадиях инновационных компаний применимы ограниченно из-за отсутствия устойчивой выручки, коротких рядов отчётности и высокой неопределённости. А.А. Долгая на примере ПАО «Яндекс» обосновывает необходимость комплексной оценки, учитывающей внутренние и внешние факторы [5]; Н.А. Никитин развивает оценку инвестиций в технологии искусственного интеллекта на основе модели жизненного цикла [9]; возможности интеллектуальных методов для интегральной оценки привлекательности на разнородных данных раскрыты Р.В. Губаревым с соавторами [4]. Для соизмерения технологической зрелости применяется семейство уровней готовности: TRL (Technology Readiness Level), изначально предложенный NASA и впоследствии дополненный уровнями инвестиционной (IRL), производственной (MRL) и коммерческой (CRL) готовности; в российской практике эти шкалы используются институтами инновационного развития.

Машинное обучение в финансовом скоринге и прогнозировании несостоятельности. Краеугольным методом прогнозирования банкротства остаётся дискриминантная модель Э. Альтмана, представляющая вероятность несостоятельности как линейную функцию финансовых коэффициентов [14]; близкую логику реализует модель Р. Таффлера, разработанная на британских данных [19]. С развитием вычислительных методов классические

дискриминантные модели всё чаще дополняются и превосходятся алгоритмами машинного обучения: как показывает Т.К. Бойкова, методы регрессионного анализа, деревьев решений и ансамблей обеспечивают более высокую точность прогнозирования банкротства предприятий [2]; близкие результаты для задач оценки кредитоспособности субъектов малого и среднего предпринимательства получены В.В. Заболоцкой [6]. Стандартом для работы с табличными финансово-экономическими данными стал градиентный бустинг над деревьями решений в реализации XGBoost, предложенной Т. Ченом и К. Гуэстрином [15]. Применительно к венчурной сфере П.Р. Юлдашев систематизирует возможности алгоритмов прогнозной аналитики (случайный лес, XGBoost, логистическая регрессия) в управлении венчурными инвестициями на этапе отбора технологических стартапов и обосновывает их преимущества для повышения объективности оценки и снижения рисков [13]. Важный для настоящей работы вывод получен Н.Е. Садковской с соавторами: экспериментально оценив применимость машинного обучения к первичному анализу успешности стартапов, авторы показали, что подобные методы наиболее эффективны именно на этапе первичной фильтрации, тогда как для финального инвестиционного решения требуется их дополнительная настройка [12]. На международном уровне S. Mashhadī с соавторами демонстрируют, что интерпретируемые ансамблевые модели позволяют прогнозировать привлечение финансирования, патентную активность и выходы стартапов с площадью под ROC-кривой 0,82–0,92 при соблюдении методической защиты от утечки данных [18].

Объяснимый искусственный интеллект и управление рисками. Практическое применение моделей машинного обучения в финансовых решениях требует интерпретируемости результатов. Универсальный подход к объяснению предсказаний на основе значений Шепли из кооперативной теории игр предложен С. Лундбергом и С.-И. Ли (метод SHAP) [17]; он позволяет количественно определять вклад каждого фактора в итоговую оценку. Методологической рамкой

для встраивания такой оценки в контур принятия инвестиционных решений служат принципы менеджмента риска, закреплённые в ГОСТ Р ИСО 31000-2019 [3], а актуальный ландшафт глобальных рисков, влияющих на инвестиционную среду, систематизирован в докладе Всемирного экономического форума [20].

Проведённый обзор показывает, что, несмотря на активное развитие отдельных направлений, в отечественной литературе отсутствует целостная объяснимая методика предынвестиционной оценки инновационных компаний, которая одновременно опиралась бы на отечественные открытые данные, интегрировала финансово-аналитический блок прогнозирования несостоятельности и подтверждала бы свою предсказательную способность на независимых форвардных исходах. Восполнению этого пробела посвящена настоящая работа, что соответствует и стратегическим ориентирам Концепции технологического развития на период до 2030 года [10].

Материалы и методы. Эмпирическая база. Исследование выполнено на массиве из 5 166 отечественных инновационных компаний, агрегированном из открытых государственных реестров и отечественных аналитических сервисов (ЕГРЮЛ, ресурс бухгалтерской (финансовой) отчётности ФНС России, базы Федерального института промышленной собственности, сервисы Chesko и Rusprofile и др.), в том числе из баз институтов инновационного развития, включая Фонд «Сколково». Массив актуален на 2025 г. Для каждой компании доступны уровни готовности (TRL, IRL, MRL, CRL), сведения о патентах, технологиях, отраслях и продуктах, а также бухгалтерская (финансовая) отчётность за 2020–2024 гг. по ИНН. Отчётность доступна для 95,8 % компаний выборки (4 950 из 5 166).

Инженерия признаков. Из исходных данных извлекается фиксированный вектор из 97 числовых признаков, объединённых в группы (табл. 1). Категориальные признаки преобразуются методом унитарного кодирования (one-hot encoding): отраслевой кластер компании (например, информационные технологии, промышленные технологии, биомедицина, энергетика) представляется

набором бинарных столбцов, в каждом из которых единица означает принадлежность к соответствующей отрасли. Наряду с базовыми показателями вектор включает блок из 58 финансовых коэффициентов из бухгалтерской отчётности: ликвидности и платёжеспособности, финансовой устойчивости, рентабельности, оборачиваемости и структуры активов, адаптированные показатели прогнозирования несостоятельности (Z-счёт Альтмана и показатель Таффлера), бинарные индикаторы риска (отрицательный собственный капитал, убыточность, повышенная долговая нагрузка) и динамику выручки и прибыли.

Таблица 1

Группы признаков методики оценки

Группа признаков	Состав	Кол-во
Финансовые	Логарифмы выручки и прибыли за 2020–2025 гг., тренд, маржа, стабильность, число лет с выручкой	18
Технологические	TRL, IRL, MRL, CRL, число патентов, число технологий, признак ИИ, число продуктов	8
Рыночные	Число отраслей, число проектов, наличие выручки, возраст компании	4
Категориальные и текстовые	Отраслевой кластер (унитарное кодирование), статус, длины описаний	9
Финансовые коэффициенты (БФО)	Ликвидность, устойчивость, рентабельность, оборачиваемость, структура активов, модели Альтмана и Таффлера, риск-флаги, динамика	58
Итого		97

Источник: составлено автором

Целевые переменные. Оценка строится по шести независимым измерениям инвестиционной привлекательности: технологическая зрелость, инновационность, рыночный потенциал, готовность команды (коммерциализация), финансовое здоровье и интегральный балл. Эталонные значения формируются прозрачной многокритериальной экспертной методикой, которая нормирует уровни готовности и финансовые показатели по шкале от 1 до 10 на основе принятых пороговых значений; модели машинного обучения обучаются воспроизводить и обобщать эту методику. Подчеркнём методологически важное обстоятельство: на первом уровне валидации целевая переменная имеет экспертно-эвристическую природу, поэтому

высокие значения метрик качества следует интерпретировать как меру внутренней согласованности и воспроизводимости методики, а не как доказательство прогноза будущего коммерческого успеха; последнее проверяется отдельно на втором уровне.

Алгоритм. В качестве базового алгоритма используется градиентный бустинг над деревьями решений (XGBoost) — последовательное обучение ансамбля решающих деревьев, в котором каждое следующее дерево корректирует ошибки предыдущих [15]. Для каждого из шести измерений обучается отдельная регрессионная модель со следующими параметрами: число деревьев — 300, максимальная глубина — 6, темп обучения — 0,05, доли подвыборки объектов и признаков — по 0,8, коэффициенты L1- и L2-регуляризации — 0,1 и 1,0 соответственно.

Математический аппарат оценки качества. Качество регрессионных моделей измеряется тремя метриками. Коэффициент детерминации R^2 показывает долю объяснённой дисперсии целевой переменной:

$$R^2 = 1 - \Sigma_i (y_i - \hat{y}_i)^2 / \Sigma_i (y_i - \bar{y})^2,$$

где y_i — фактическое значение, $\hat{y}_i$ — прогноз модели, $\bar{y}$ — среднее по выборке, n — число объектов. Значение $R^2 = 1$ соответствует идеальному прогнозу, $R^2 = 0$ — модели, эквивалентной среднему. Средняя абсолютная ошибка (MAE) и корень из средней квадратичной ошибки (RMSE) измеряют величину отклонения прогноза в баллах шкалы:

$$MAE = \frac{1}{n} \cdot \Sigma_i |y_i - \hat{y}_i|; \quad RMSE = \sqrt{\frac{1}{n} \cdot \Sigma_i (y_i - \hat{y}_i)^2}.$$

Для задачи бинарной классификации (прогноз направления изменения выручки) используется площадь под ROC-кривой (ROC-AUC) — вероятность того, что случайно выбранному положительному объекту модель присвоит более

высокую оценку, чем случайно выбранному отрицательному; значение 0,5 соответствует случайному угадыванию, 1,0 — идеальному разделению.

Финансовые модели прогнозирования несостоятельности. Z-счёт Альтмана рассчитывается как взвешенная сумма пяти финансовых коэффициентов:

$$Z = 1,2 \cdot X_1 + 1,4 \cdot X_2 + 3,3 \cdot X_3 + 0,6 \cdot X_4 + 1,0 \cdot X_5,$$

где X_1 — отношение чистого оборотного капитала к активам, X_2 — нераспределённая прибыль к активам, X_3 — операционная прибыль к активам, X_4 — собственный капитал к обязательствам, X_5 — выручка к активам. Значения $Z > 2,99$ трактуются как зона финансовой устойчивости, $Z < 1,81$ — как зона высокого риска несостоятельности. Показатель Таффлера рассчитывается по формуле:

$$Z_T = 0,53 \cdot T_1 + 0,13 \cdot T_2 + 0,18 \cdot T_3 + 0,16 \cdot T_4,$$

где T_1 — операционная прибыль к краткосрочным обязательствам, T_2 — оборотные активы к суммарным обязательствам, T_3 — краткосрочные обязательства к активам, T_4 — выручка к активам; значение $Z_T > 0,3$ указывает на низкий риск банкротства.

Объяснимость. Для содержательной интерпретации применяется метод SHAP, определяющий вклад φ_j каждого фактора j в итоговую оценку на основе значений Шепли из кооперативной теории игр [17]:

$$\varphi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} [f_{(S \cup \{j\})}(x_{(S \cup \{j\})}) - f_S(x_S)],$$

где F — множество всех признаков, S — произвольное подмножество без признака j , $f(S)$ — прогноз модели на подмножестве S . Итоговая оценка раскладывается на базовое значение и сумму вкладов: $\hat{y} = \varphi_0 + \sum_j \varphi_j$. Например, для условной компании при базовом значении $\varphi_0 = 5,0$ вклады IRL (+0,9), TRL (+0,6) и роста выручки (+0,4) при отрицательном вкладе отсутствия патентов (−0,3) дают итоговую оценку 6,6 балла, что и предъявляется инвестору в виде объяснения.

Двухуровневая валидация. Первый уровень — оценка воспроизведения экспертной методики методом пятиблочной перекрёстной проверки (k-fold cross-validation, $k = 5$), усреднённой по трём значениям инициализации генератора случайных чисел. Второй уровень — независимая форвардная валидация, исключая «протечку» разметки: на признаках, рассчитанных только по данным 2020–2023 гг., прогнозируется финансовый исход 2024 г. — логарифм выручки (регрессия) и направление её изменения относительно предыдущего года (бинарная классификация). Дополнительно индикаторы Альтмана и Таффлера сопоставляются с фактическим статусом компаний (действующая либо находящаяся в процедуре ликвидации/банкротства), а значимость различий между группами оценивается непараметрическим критерием Манна–Уитни.

Сравнение и реализация. Качество градиентного бустинга сопоставляется с базовыми алгоритмами (линейная регрессия, случайный лес); проводится абляционный анализ вклада финансового блока. Программная реализация выполнена на открытом стеке (Python, библиотеки xgboost, scikit-learn, shap), не зависящем от зарубежных облачных сервисов. Методика доведена автором до действующего облачного аналитического сервиса (прототип AutoScoutBot), что подтверждает её практическую применимость.

Параметрическая модель экономического эффекта. Экономический эффект автоматизации измеряется как сокращение совокупных издержек и времени первичного скрининга пула из N проектов. Для ручного подхода совокупные издержки составляют $C_{\text{ручн}} = N \cdot c_m$, время — $T_{\text{ручн}} = N \cdot t_m$, где c_m — стоимость экспертизы одного проекта, t_m — её длительность. Для автоматизированного подхода совокупные издержки складываются из фиксированных затрат C_0 (разработка, инфраструктура, доступ к данным) и предельной стоимости оценки одного проекта c_a (обращения к API открытых данных и вычисления): $C_{\text{авто}} = C_0 + N \cdot c_a$, причём $c_a \ll c_m$; время $T_{\text{авто}} = N \cdot t_a$, где t_a измеряется секундами. Относительная экономия издержек и времени рассчитывается как $\eta_c = 1 - C_{\text{авто}} /$

$C_{ручн}$ и $\eta_t = 1 - T_{авто} / T_{ручн}$, а точка окупаемости автоматизации (число проектов, при котором её совокупные издержки сравниваются с ручным подходом) — как $N^* = C_0 / (c_m - c_a)$. Сопоставление корректно для функции первичного скрининга: автоматизированная оценка не заменяет углублённый due diligence отобранных проектов, а замещает дорогостоящий ручной первичный отбор всего пула.

Результаты. Качество многомерной модели. Результаты пятиблочной перекрёстной проверки, усреднённой по трём инициализациям, представлены в табл. 2. Среднее по шести измерениям значение R^2 составляет 0,968 при низком стандартном отклонении (около 0,008), что свидетельствует об устойчивости результата. Модели воспроизводят экспертную методику с высокой точностью: средняя абсолютная ошибка не превышает 0,15 балла по десятибалльной шкале.

Таблица 2

Качество моделей по шести измерениям (XGBoost, пятиблочная кросс-валидация, усреднение по трём инициализациям)

Измерение	R^2 (среднее $\pm$ ст. откл.)	MAE	RMSE
Интегральный балл	$0,967 \pm 0,008$	0,143	0,322
Технологическая зрелость	$0,966 \pm 0,009$	0,142	0,440
Инновационность	$0,973 \pm 0,007$	0,113	0,354
Рыночный потенциал	$0,969 \pm 0,009$	0,113	0,366
Готовность команды	$0,966 \pm 0,008$	0,087	0,290
Финансовое здоровье	$0,970 \pm 0,008$	0,142	0,398
Среднее	0,968		

Источник: расчёты автора

Сравнение алгоритмов. Градиентный бустинг превосходит базовые алгоритмы как по коэффициенту детерминации, так и по средней абсолютной ошибке (табл. 3), что подтверждает обоснованность выбора метода для работы с разнородными табличными данными.

Таблица 3

Сравнение алгоритмов (интегральный балл, пятиблочная кросс-валидация)

Алгоритм	R^2	MAE
Линейная регрессия	0,953	0,230
Случайный лес	0,951	0,242
Градиентный бустинг (XGBoost)	0,967	0,143

Источник: расчёты автора

Факторы оценки. Анализ методом SHAP (рис. 1) показывает, что наибольший вклад в интегральную оценку вносят уровни готовности технологии (IRL и TRL), масштаб выручки, уровень коммерческой и производственной готовности (CRL, MRL), а также патентная активность и признак применения технологий искусственного интеллекта. Полученное распределение экономически интерпретируемо: оценка инвестиционной привлекательности определяется сочетанием технологической зрелости и финансовой состоятельности компании, что согласуется с практикой экспертного отбора.

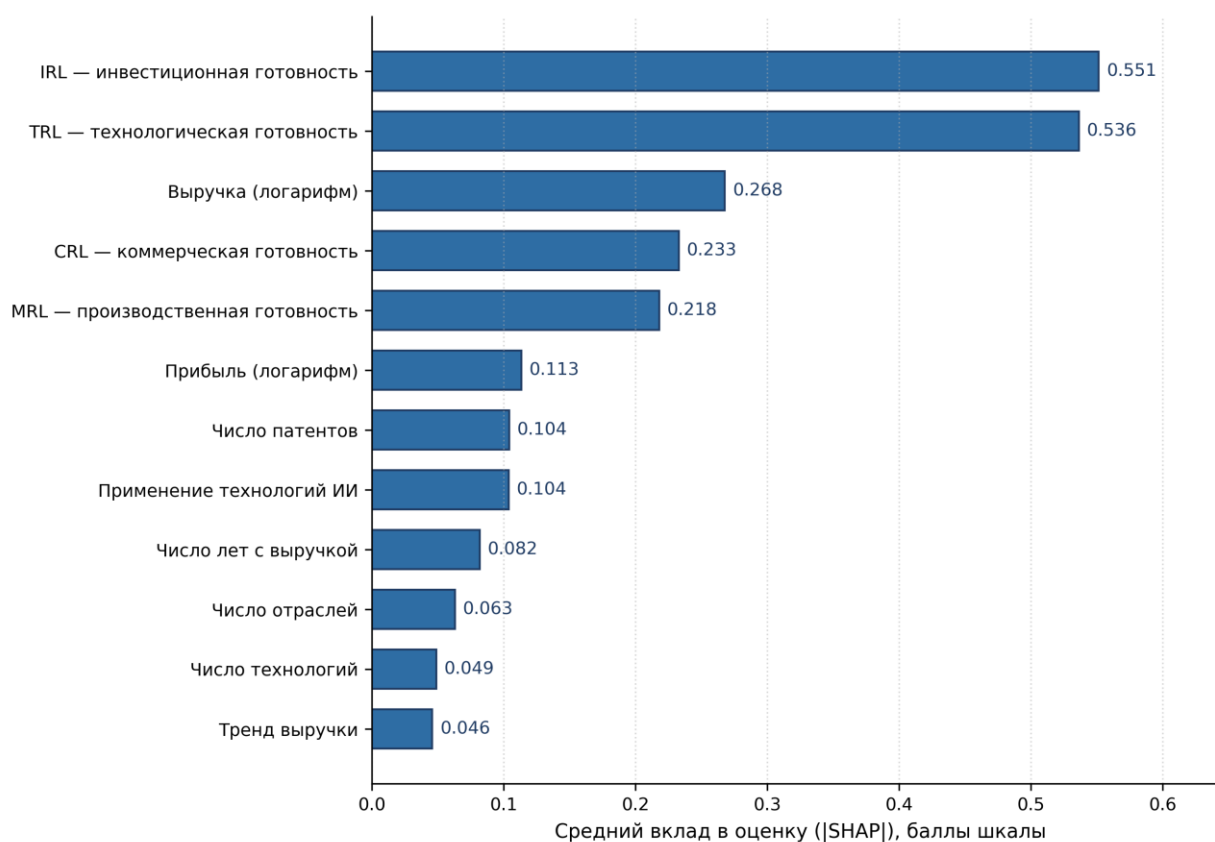


Рис. 1 – Вклад факторов в интегральную оценку по методу SHAP (топ-факторы: IRL, TRL, выручка, CRL, MRL, патенты) (составлен автором)

Независимая форвардная валидация. На задаче прогноза логарифма выручки 2024 г. по данным 2020–2023 гг. (выборка 3 181 компания, доля проверки — 20 %) достигнут коэффициент детерминации $R^2 = 0,77$ при средней абсолютной ошибке 0,74, что свидетельствует о реальной предсказательной способности

модели в отношении будущей финансовой траектории. На более сложной задаче прогноза направления изменения выручки площадь под ROC-кривой составила 0,62 (доля растущих компаний в выборке — 0,69), что соответствует умеренной, но значимой различительной способности и отражает объективную трудность прогнозирования краткосрочной динамики на открытых данных.

Валидация финансового блока. Сопоставление интегральных индикаторов несостоятельности с фактическим статусом компаний подтверждает содержательность финансового блока (табл. 4): у действующих компаний медианный Z-счёт Альтмана составляет 1,42, тогда как у компаний в процедуре ликвидации или банкротства — 0,60; аналогичное разделение наблюдается по показателю Таффлера. Различие между группами статистически значимо по непараметрическому критерию Манна–Уитни ($p < 0,001$ для Z-счёта Альтмана и $p < 0,05$ для показателя Таффлера). Несмотря на малую долю проблемных компаний в выборке (41 из 5 166), различие распределений устойчиво и согласуется с экономическим смыслом моделей.

Таблица 4

Индикаторы несостоятельности по группам компаний

Группа	Кол-во	Z-счёт Альтмана (медиана)	Показатель Таффлера (среднее)
Действующие	4 909	1,42	3,19
В процедуре ликвидации/банкротства	41	0,60	1,48

Источник: расчёты автора

Абляционный анализ показал, что исключение блока из 58 финансовых коэффициентов практически не влияет на воспроизведение интегрального экспертного балла (R^2 изменяется в пределах статистической погрешности: 0,967 против 0,968). Этот результат закономерен: эталонная разметка интегрального балла опирается прежде всего на уровни готовности и масштаб выручки. Вместе с тем, как показывает табл. 4, финансовый блок несёт самостоятельную ценность

именно как независимый контур оценки риска несостоятельности, валидируемый на реальных исходах, а не на экспертной разметке.

Сравнительная характеристика рыночных решений. Сопоставление разрабатываемой методики с представленными на рынке аналитическими решениями (таблица 5) показывает, что ни одно из них не сочетает одновременно автоматизированную оценку на основе машинного обучения, объяснимость результатов и опору на отечественные открытые данные при низкой стоимости доступа.

Таблица 5

Сравнительная характеристика аналитических решений венчурного рынка

Параметр	Справочные системы (СПАРК, Контур.Фокус)	Зарубежные платформы скаутинга (Harmonic.ai, PitchBook)	Консалтинг (due diligence)	Предлагаемая методика
Тип решения	Справочная база данных	Скаутинг/сорсинг сделок	Ручная экспертиза	Облачный сервис
Основная технология	Агрегация реестров	Машинное обучение, сигналы	Экспертный анализ	Машинное обучение
Охват данных РФ	Полный	Слабый	Полный	Полный
Автоматическая оценка	Нет	Частично	Нет	Да (6 измерений)
Объяснимость	Нет	Ограниченно	Экспертное заключение	Да (SHAP)
Ориентир стоимости	32–99 тыс. руб./год; ~25 тыс. руб./мес [8]	от ~12–25 тыс. долл./год [16]	50–500 тыс. руб.+ за проект [1]	Низкая; предельная стоимость ≈ 2 руб./проект

Источник: составлено автором по данным [1; 8; 16]

Оценка экономического эффекта. Параметры модели оценены по проектной смете прототипа. Разовые затраты складываются из трудоёмкости разработки и инфраструктуры. Ожидаемая трудоёмкость рассчитана вероятностным методом (PERT) и составила 238,8 чел.-ч.; стоимость часа принята равной 1000 руб./ч. как консервативная нижняя граница рыночной ставки разработки (медианной заработной плате ИТ-специалиста соответствует около

1400 руб./ч.). С учётом инфраструктурно-сервисных расходов (хостинг, домен, доступ к данным — порядка 7,5 тыс. руб.) фиксированные затраты составляют $C_0 = 238,8 \cdot 1000 + 7\,500 \approx 0,25$ млн руб. Предельная стоимость одного прогона определяется несколькими обращениями к API открытых данных (порядка 0,15 руб. за запрос) и вычислениями и не превышает $c_a \approx 2$ руб. Ручная же экспертиза одного проекта стоит $c_m = 120\text{--}500$ тыс. руб. и занимает дни–недели [1]. Подставляя эти значения в модель, получаем совокупные издержки первичного скрининга пула из 100 проектов: 12–50 млн руб. при ручном подходе против $\approx 0,25$ млн руб. при автоматизированном (табл. 6); разовые вложения окупаются уже на первом–втором проекте ($N^* = C_0 / (c_m - c_a) < 2$). Экономический смысл эффекта — не «бесплатность» оценки, а перевод затрат из переменных (труд эксперта на каждый проект) в преимущественно постоянные (разовая разработка), благодаря чему первичный скрининг перестаёт быть «узким местом» и масштабируется на тысячи компаний. При этом автоматизированная оценка не заменяет углублённый due diligence отобранных проектов, а замещает ручной первичный отбор всего пула.

Таблица 6

Сравнительная экономика первичного скрининга (пул $N = 100$ проектов)

Показатель	Ручная экспертиза	Автоматизированная оценка
Разовые вложения (разработка, инфраструктура), млн руб.	—	$\approx 0,25$
Предельная стоимость оценки одного проекта	120–500 тыс. руб.	≈ 2 руб.
Время оценки одного проекта	дни–недели	секунды
Совокупные издержки на пул 100 проектов, млн руб.	12–50	$\approx 0,25$
Точка окупаемости вложений N^*	—	1–2 проекта

Источник: расчёты автора; c_m и сроки — по данным [1]; C_0 – по смете трудоёмкости проекта (метод PERT, 238,8 чел.-ч.).



Рис. 2 – Архитектура методики автоматизированной оценки (запрос → семантический поиск → многомерный скоринг → объяснение результата) (составлен автором)

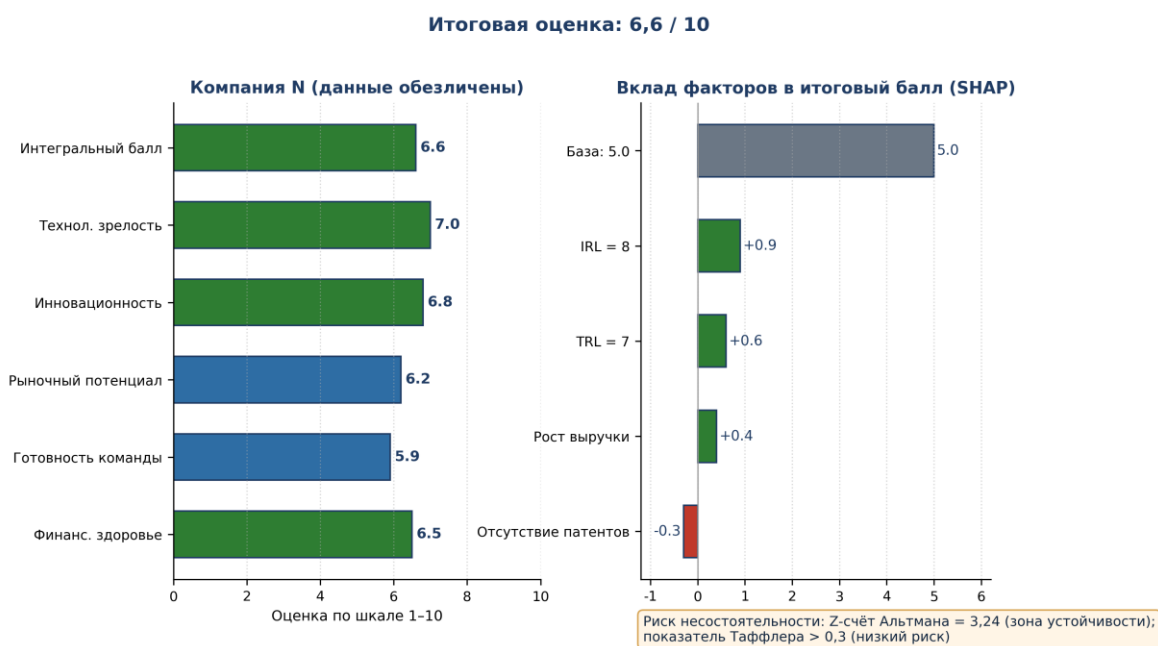


Рис. 3 – Пример объяснимой карточки оценки компании (данные обезличены) (составлен автором)

Обсуждение. Полученные результаты позволяют сформулировать несколько обобщающих наблюдений.

Во-первых, предложенная двухуровневая схема валидации снимает ключевое методологическое возражение к автоматизированным скоринговым системам. Высокое значение R^2 на первом уровне (0,968) корректно интерпретировать не как «точность прогноза успеха», а как меру воспроизводимости прозрачной экспертной методики: модель надёжно тиражирует логику эксперта, что ценно для масштабирования, но не доказывает предсказательную силу. Последняя подтверждается на втором, независимом уровне — прогнозом будущей выручки с $R^2 = 0,77$ в условиях, исключающих информационную «протечку». Такое разделение «согласованности» и «прогноза» отличает работу от коммерческих решений, метрики которых, как правило, не раскрываются.

Во-вторых, финансово-аналитический блок выполняет в методике особую роль. Абляционный анализ показал, что для воспроизведения интегрального экспертного балла он избыточен, однако именно он обеспечивает независимый контур оценки риска несостоятельности: индикаторы Альтмана и Таффлера статистически разделяют действующие и проблемные компании на фактических исходах. Тем самым методика сочетает оценку потенциала (через уровни готовности и рыночные показатели) с оценкой риска (через финансовые коэффициенты), что соответствует принципам менеджмента риска [3] и практике финансового анализа.

В-третьих, полученные результаты согласуются с международными исследованиями и российской литературой. Достигнутые показатели прогноза сопоставимы по порядку величины с результатами интерпретируемых ансамблевых моделей для предсказания исходов стартапов [18], а обоснованность применения градиентного бустинга и прогнозной аналитики для отбора технологических проектов подтверждается работами П.Р. Юлдашева [13] и Т.К. Бойковой [2]. Умеренное значение ROC-AUC (0,62) на задаче прогноза краткосрочной динамики выручки согласуется с выводом Н.Е. Садковской с соавторами о том, что машинное обучение наиболее результативно именно на этапе первичной фильтрации проектов

[12], что и определяет позиционирование предлагаемой методики. Содержательный вывод о ведущей роли технологической зрелости и финансовой состоятельности в оценке привлекательности согласуется с выводами о необходимости комплексного учёта внутренних и внешних факторов [4; 5; 9].

В-четвёртых, в отличие от зарубежных платформ венчурной аналитики, распространяемых по корпоративным тарифам и слабо охватывающих российскую юрисдикцию [16], предлагаемая методика реализована на отечественных открытых данных и открытом программном стеке, что отвечает приоритетам Концепции технологического развития до 2030 года [10] и повышает устойчивость аналитического контура к внешним рискам [20]. Её практическая значимость определяется снижением транзакционных издержек первичного отбора примерно на два порядка и демократизацией доступа к аналитике.

Исследование имеет ряд ограничений. Эталонная разметка интегрального балла носит экспертно-эвристический характер; число компаний с фактически наблюдаемым неблагоприятным исходом невелико, что ограничивает статистическую мощность валидации финансового блока; выборка, хотя и масштабна, отражает определённый контур инновационных компаний. Перспективными направлениями являются расширение охвата за счёт компаний иных экосистем, обогащение данных из патентных баз и реестров государственных контрактов, а также накопление наблюдаемых исходов для прямой калибровки моделей на реальном коммерческом успехе.

Заключение. В работе разработана и эмпирически валидирована методика многомерной объяснимой предынвестиционной оценки инновационных компаний на основе машинного обучения и открытых данных. Методика объединяет шесть измерений инвестиционной привлекательности с финансово-аналитическим блоком, включающим адаптированные модели прогнозирования несостоятельности Альтмана и Таффлера. На массиве более 5 тыс. отечественных компаний показано, что модели воспроизводят экспертную методику со средним R^2

= 0,968, превосходят базовые алгоритмы, а на независимой форвардной задаче прогноза выручки достигают $R^2 = 0,77$; финансовые индикаторы статистически разделяют действующие и проблемные компании. Параметрическая модель показывает, что автоматизация снижает совокупные издержки первичного скрининга на порядки при окупаемости разовых вложений уже на единицах проектов.

Научная новизна работы состоит в авторской методике многомерного объяснимого скоринга с интегрированным блоком финансовых рисков, в двухуровневой схеме валидации, отделяющей внутреннюю согласованность от предсказательной способности, и в параметрической оценке экономического эффекта автоматизации. Практическая ценность заключается в обеспечении доступного и масштабируемого инструмента первичного отбора инновационных проектов для широкого круга инвесторов. Дальнейшая работа будет направлена на расширение эмпирической базы и прямую калибровку моделей на наблюдаемых исходах.

Библиографический список:

1. Аллент-Аудит. Due Diligence компании: проверка рисков, сроки и стоимость [Электронный ресурс]. – URL: <https://www.allent-audit.ru/services/audit/proverka-biznesa-due-diligence/> (дата обращения: 16.06.2026).
2. Бойкова Т.К. Использование машинного обучения для прогнозирования банкротства предприятий // Актуальные исследования. – 2024. – № 12 (194), ч. II. – С. 35–37.
3. ГОСТ Р ИСО 31000-2019. Менеджмент риска. Принципы и руководство. – М.: Стандартиформ, 2020. – 19 с.
4. Губарев Р.В., Дзюба Е.И., Хазиев Р.С. Оценка инвестиционной привлекательности субъектов РФ с применением искусственного интеллекта // Финансы: теория и практика. – 2025. – Т. 29, № 2. – С. 120–136.

5. Долгая А.А. Инвестиционный анализ IT-компаний: исследование на примере Яндекса // Review of Business and Economics Studies. – 2022. – Т. 10, № 3. – С. 33–55.
6. Заболоцкая В.В. Методы диагностики и прогнозирования кредитоспособности субъектов МСП с применением искусственного интеллекта // Известия Саратовского университета. Новая серия. Серия: Экономика. Управление. Право. – 2024. – Т. 24, вып. 3. – С. 294–311.
7. Исследование российского венчурного рынка 2025 года [Электронный ресурс] / Русский Венчур, Malina VC, Brio Capital // vc.ru. – URL: <https://vc.ru/id4994551/2652053-issledovanie-rossijskogo-venchurnogo-rynka-2025-goda> (дата обращения: 16.06.2026).
8. Контур.Фокус. Тарифы и стоимость подписки [Электронный ресурс]. – URL: <https://kontur.ru/focus/prices> (дата обращения: 16.06.2026).
9. Никитин Н.А. Применение жизненного цикла модели для оценки инвестиций в искусственный интеллект на примере больших языковых моделей // Финансы: теория и практика. – 2024. – Т. 28, № 3. – С. 206–217.
10. Распоряжение Правительства Российской Федерации от 20.05.2023 № 1315-р «Об утверждении Концепции технологического развития на период до 2030 года». – URL: <http://publication.pravo.gov.ru/Document/View/0001202305250050> (дата обращения: 16.06.2026).
11. Рынок венчурных инвестиций России. I полугодие 2025: отчёт Московского инновационного кластера [Электронный ресурс] // Коммерсантъ. – URL: <https://www.kommersant.ru/doc/7943717> (дата обращения: 16.06.2026).
12. Садковская Н.Е., Манин А.Н., Горшкова А.П., Прудников А.В. Практическое применение методов машинного обучения и искусственного интеллекта при первичном анализе успешности стартапов // Нейрокомпьютеры: разработка, применение. – 2025. – Т. 27, № 2. – С. 13–22.
13. Юлдашев П.Р. Применение алгоритмов машинного обучения в управлении венчурными инвестициями на этапе отбора технологических стартапов // Экономика, предпринимательство и право. – 2025. – Т. 15, № 8. – С. 5247–5264.

14. Altman E.I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy // The Journal of Finance. – 1968. – Vol. 23, № 4. – P. 589–609.
15. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. – 2016. – P. 785–794.
16. Harmonic.ai. The startup discovery engine [Электронный ресурс]. – URL: <https://harmonic.ai> (дата обращения: 16.06.2026).
17. Lundberg S.M., Lee S.-I. A Unified Approach to Interpreting Model Predictions // Advances in Neural Information Processing Systems 30 (NIPS 2017). – 2017. – P. 4765–4774.
18. Mashhadi S., Saghezchi A., Ghassemzadeh Kashani V. Interpretable Machine Learning for Predicting Startup Funding, Patenting, and Exits [Электронный ресурс] // arXiv preprint arXiv:2510.09465. – 2025. – URL: <https://arxiv.org/abs/2510.09465> (дата обращения: 16.06.2026).
19. Taffler R.J. The Assessment of Company Solvency and Performance Using a Statistical Model // Accounting and Business Research. – 1983. – Vol. 13, № 52. – P. 295–308.
20. The Global Risks Report 2025. – Geneva: World Economic Forum, 2025. – 107 p.

References:

1. Allent-Audit. Due Diligence of a company: risk assessment, timing and cost. Available at: <https://www.allent-audit.ru/services/audit/proverka-biznesa-due-diligence/> (accessed: 16.06.2026).
2. Boykova T.K. (2024) Using machine learning to predict corporate bankruptcy. Aktual'nye issledovaniya, № 12 (194), part II, pp. 35–37.
3. GOST R ISO 31000-2019. Risk management. Principles and guidelines. Moscow, Standartinform, 2020. 19 p.

4. Gubarev R.V., Dzyuba E.I., Khaziev R.S. (2025) Assessment of Investment Attractiveness of RF Entities Using Artificial Intelligence. *Finance: Theory and Practice*, vol. 29, № 2, pp. 120–136.
5. Dolgaya A.A. (2022) Investment analysis of IT companies: a case study of Yandex. *Review of Business and Economics Studies*, vol. 10, № 3, pp. 33–55.
6. Zabolotskaya V.V. (2024) Methods for diagnosing and forecasting the creditworthiness of SMEs using artificial intelligence. *Izvestiya of Saratov University. Economics. Management. Law*, vol. 24, iss. 3, pp. 294–311.
7. Russian Venture Market Study 2025. Russkiy Venchur, Malina VC, Brio Capital. vc.ru. Available at: <https://vc.ru/id4994551/2652053-issledovanie-rossijskogo-venchurnogo-rynka-2025-goda> (accessed: 16.06.2026).
8. Kontur.Focus. Subscription plans and prices. Available at: <https://kontur.ru/focus/prices> (accessed: 16.06.2026).
9. Nikitin N.A. (2024) Applying a model life cycle to assess investments in artificial intelligence: the case of large language models. *Finance: Theory and Practice*, vol. 28, № 3, pp. 206–217.
10. Order of the Government of the Russian Federation dated 20.05.2023 № 1315-r «On approval of the Concept of Technological Development for the period up to 2030». Available at: <http://publication.pravo.gov.ru/Document/View/0001202305250050> (accessed: 16.06.2026).
11. Russian Venture Investment Market. H1 2025: Moscow Innovation Cluster report. Kommersant. Available at: <https://www.kommersant.ru/doc/7943717> (accessed: 16.06.2026).
12. Sadkovskaya N.E., Manin A.N., Gorshkova A.P., Prudnikov A.V. (2025) Practical application of machine learning and artificial intelligence methods in the primary analysis of startup success. *Neurocomputers: Development, Application*, vol. 27, № 2, pp. 13–22.
13. Yuldashev P.R. (2025) Application of machine learning algorithms in venture investment management at the technology startup selection stage. *Economics, Entrepreneurship and Law*, vol. 15, № 8, pp. 5247–5264.

14. Altman E.I. (1968) Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. The Journal of Finance, vol. 23, № 4, pp. 589–609.
15. Chen T., Guestrin C. (2016) XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794.
16. Harmonic.ai. The startup discovery engine. Available at: <https://harmonic.ai> (accessed: 16.06.2026).
17. Lundberg S.M., Lee S.-I. (2017) A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems 30 (NIPS 2017), pp. 4765–4774.
18. Mashhadi S., Saghezchi A., Ghassemzadeh Kashani V. (2025) Interpretable Machine Learning for Predicting Startup Funding, Patenting, and Exits. arXiv preprint arXiv:2510.09465. Available at: <https://arxiv.org/abs/2510.09465> (accessed: 16.06.2026).
19. Taffler R.J. (1983) The Assessment of Company Solvency and Performance Using a Statistical Model. Accounting and Business Research, vol. 13, № 52, pp. 295–308.
20. The Global Risks Report 2025. Geneva, World Economic Forum, 2025. 107 p.