Научная статья
УДК 519.862.6
EDN EMVKTB
DOI 10.17150/2713-1734.2026.8(3).440-461

М.П. Базилевский
Иркутский государственный университет путей сообщения,
г. Иркутск, Российская Федерация

Построение регрессионных моделей как система: выявление факторов интерпретируемости и её эмерджентный характер

Аннотация. Проблема интерпретируемости моделей машинного обучения в настоящее время относится к числу наиболее актуальных направлений научных исследований. Регрессионный анализ, будучи одним из классических инструментов данной области, также сталкивается с этой проблемой. При этом остаётся открытым вопрос о возможности целенаправленного управления степенью интерпретируемости регрессионных моделей. Цель настоящей работы — доказать с позиций системного анализа, что процесс построения регрессионных моделей это система, а также выявить элементы системы, которые оказывают наиболее сильное влияние на интерпретируемость результирующей модели. Процесс построения регрессионных моделей рассмотрен как целенаправленная, структурированная и иерархическая система. Выделено 7 подсистем системы. Установлено, что процесс построения регрессионных моделей является открытой системой, обменивающейся с внешней средой информацией и ресурсами. У системы зафиксировано наличие чётких входов и выходов, а также выявлен контур обратной связи. На примере клинической эндокринологии и промышленного машиностроения продемонстрировано, что использование неинтерпретируемой результирующей регрессионной зависимости в реальной практике для принятия управленческих решений способно приводить к негативным, вплоть до катастрофических, последствиям. Показано, что интерпретируемость носит эмерджентный характер. Иными словами, она не сводится к сумме свойств отдельных этапов построения модели и не может быть гарантирована качественным выполнением каждого из них в отдельности, а возникает исключительно на системном уровне. В рамках системного подхода проведена декомпозиция каждого из первых шести этапов (подсистем) на 56 элементарных составляющих (операций). Из них выбрано 23 операции, наиболее сильно влияющих на интерпретируемость регрессионной модели.

Ключевые слова. Регрессионный анализ, модель, интерпретируемость, системный анализ, внешняя среда, обратная связь, свойство эмерджентности.

Информация о статье. Дата поступления 5 августа 2026 г.; дата принятия к публикации: 13 октября 2026 г.; дата онлайн-размещения: 11 ноября 2026 г.

Для цитирования. Базилевский М.П. Построение регрессионных моделей как система: выявление факторов интерпретируемости и её эмерджентный характер / М.П. Базилевский. — DOI 10.17150/2713-1734.2026.8(3).440-461. — EDN EMVKTB // System Analysis & Mathematical Modeling. — 2026. — Т. 8, № 3. — С. 440–461.

© Базилевский М.П., 2026

Original article

M.P. Bazilevskiy
Irkutsk State Transport University,
Irkutsk, Russian Federation

Regression Modeling as a System: Identifying Factors of Interpretability and Its Emergent Nature

Abstract. The problem of interpretability of machine learning models currently ranks among the most pressing areas of scientific research. Regression analysis, being one of the classical tools within this domain, also faces this challenge. At the same time, the question of whether the degree of interpretability of regression models can be deliberately controlled remains open. The aim of this study is to demonstrate, from the standpoint of systems analysis, that the process of constructing regression models constitutes a system, and to identify the elements of this system that exert the strongest influence on the interpretability of the resulting model. The process of constructing regression model is considered as a goal-oriented, structured, and hierarchical system. Seven subsystems are distinguished within it. It is established that the process of constructing regression models is an open system that exchanges information and resources with the external environment. The system exhibits clearly defined inputs and outputs, and a feedback loop is identified. Using examples from clinical endocrinology and industrial mechanical engineering, it is demonstrated that the application of non-interpretable resulting regression relationships in real-world practice for managerial decision-making can lead to adverse, even catastrophic, consequences. It is shown that interpretability is emergent in nature. In other words, it is not reducible to the sum of the properties of individual stages of model construction and cannot be guaranteed by the high-quality execution of each of them in isolation; rather, it arises exclusively at the system level. Within the framework of the systems approach, a decomposition of each of the first six stages (subsystems) into 56 elementary components (operations) is performed. Of these, 23 operations that most strongly affect the interpretability of the regression model are selected.

Keywords. Regression analysis, model, interpretability, systems analysis, external environment, feedback, emergent property.

Article info. Received 5 August, 2026; Accepted 10 October, 2026; Available online 11 November, 2026.

For Citation. Bazilevskiy M.P. Regression Modeling as a System: Identifying Factors of Interpretability and Its Emergent Nature. *System Analysis & Mathematical Modeling*, 2026, vol. 8, no. 3, pp. 440–461. (In Russian). EDN: EMVKTB. DOI: 10.17150/2713-1734.2026.8(3).440-461.

Введение

В настоящее время в условиях стремительной интеграции моделей машинного обучения в критически важные сферы, такие, как, например, медицина и финансы, проблема их «черного ящика» становится ключевым барьером для внедрения. Поэтому продолжает развиваться такое научное направление, как интерпретируемое машинное обучение [1; 2], которое предлагает методологическую базу для обеспечения прозрачности алгоритмических решений. Именно это свойство делает его развитие необходимым условием для построения доверительных и этичных систем искусственного интеллекта.

На основе анализа литературных источников [3–7] можно заключить, что область интерпретируемого машинного обучения демонстрирует активное развитие как в методологическом, так и в прикладном аспектах. В частности, проведенные в [3] масштабные исследования подтверждают, что современные обобщенные аддитивные модели достигают прогностической точности, сопоставимой с лучшими «черными ящиками», сохраняя при этом структурную прозрачность. Сравнительный анализ в [4] на разнородных данных показывает, что, несмотря на лидерство «черных ящиков» в чистой точности, классические статистические модели сохраняют стратегическое преимущество там, где важны прозрачность, статистический вывод и доверие пользователя. Помимо фундаментальных сравнений, активно развиваются и конкретные практические реализации. Так, в работе [5] представлена интерпретируемая модель для пространственного картирования температуры поверхности земли в городской среде Дананга (Вьетнам), в [6] строятся интерпретируемые модели прогнозирования смертности и длительности пребывания в реанимации, а в [7] разработана интерпретируемая модель для прогнозирования скорости ветра в прибрежной зоне Abbot Point (Австралия). Таким образом, совокупность представленных работ свидетельствует о высокой востребованности интерпретируемых методов как универсального инструмента для задач, где баланс между точностью и обоснованностью решения является критическим условием практического внедрения.

В арсенал методов машинного обучения традиционно входит регрессионный анализ. Вместе с тем в научной среде зачастую априорно постулируется интерпретируемость регрессионных моделей. Так, в работе [8] утверждается, что «обучаемые модели являются интерпретируемыми, если они выражаются в виде математических уравнений», что фактически приравнивает любую параметрическую регрессионную модель к классу интерпретируемых. Однако подобная позиция представляется дискуссионной. Альтернативный подход, разделяемый автором настоящей работы, рассмотрен в [9], где опровергается распространённое заблуждение о том, что линейные регрессионные модели являются полностью интерпретируемыми. Авторы [9] обосновывают, что на практике даже линейные регрессии сопряжены с рядом серьёзных ограничений, осложняющих их интерпретацию: нарушение допущения о линейности, мультиколлинеарность и пр.

Однако методика построения интерпретируемых регрессионных моделей имеет достаточно продолжительную историю развития. Так, ещё в 2011 году была опубликована монография [10], в которой предложен алгоритм конструирования таких моделей. В настоящее время активно развивается технология построения вполне интерпретируемых регрессионных моделей [11–14]. Вме-

сте с тем остаётся открытым вопрос об управлении степенью интерпретируемости регрессионных моделей. В связи с этим в настоящей работе ставится цель провести системный анализ процесса построения регрессий. Для достижения поставленной цели необходимо решить две задачи: 1) рассмотреть процесс построения регрессионных моделей как систему с позиций системного анализа; 2) идентифицировать элементы данной системы, оказывающие наиболее существенное влияние на интерпретируемость результирующей модели.

Теоретическую основу настоящей статьи составляют положения, разработанные крупными специалистами по системному анализу — С.Л. Оптнером [15], С. Янгом [16], Н.П. Федоренко [17], Ю.И. Черняком [18], Ф.И. Перегудовым [19]. Этапы регрессионного моделирования сформулированы в соответствии с концепцией, представленной в совместном труде С.А. Айвазяна и В.С. Мхитаряна [20].

1. Процесс построения регрессионных моделей как система

Схема, доказывающая системность процесса построения регрессионных моделей, представлена на рис. 1. Рассмотрим подробнее составляющие этой схемы.

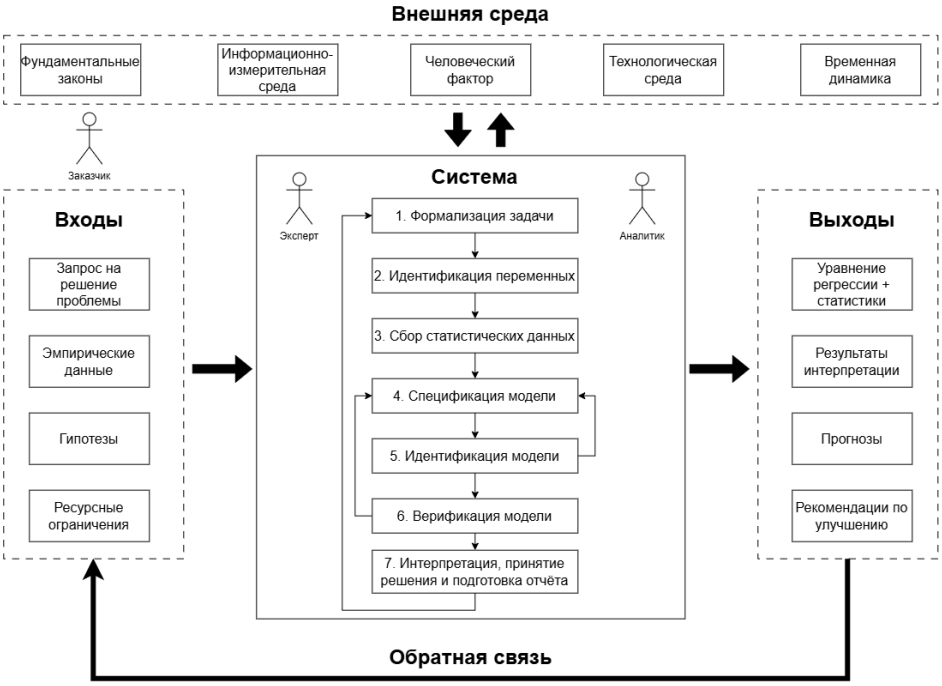


Рис. 1. Системность процесса построения регрессионных моделей

Схема на рис. 1 демонстрирует наличие в системе следующих семи *подсистем (этапов)*.

1. Подсистема «Формализация задачи». В этой подсистеме осуществляется комплексная постановка прикладной и вычислительной проблемы, включающая идентификацию запросов заказчика и содержательную интерпретацию целевого показателя. Определяются тип решаемой задачи (прогнозирование, объяснение или управление), допустимый уровень ошибки прогноза, а также приоритет между точностью, интерпретируемостью и алгоритмической простотой модели. Уточняется структура доступных данных. Одновременно с этим фиксируются ограничения по вычислительным ресурсам и времени исполнения. Особое внимание уделяется требованиям к объяснимости получаемых результатов, порогу статистической значимости оценок, необходимости построения доверительных интервалов для коэффициентов и прогнозных значений, а также формату итогового отчёта, что в совокупности обеспечивает воспроизводимость, обоснованность и практическую применимость модели.

2. Подсистема «Идентификация переменных». В этой подсистеме осуществляется формирование расширенного теоретического множества потенциальных предикторов, обоснованного содержательными гипотезами и предшествующими исследованиями. Проводится классификация отобранных объясняющих переменных по измерительным шкалам. Для временных рядов дополнительно выявляются лаговые структуры, отражающие инерционность процессов. В целях коррекции эндогенности специфицируются инструментальные переменные, а при недоступности прямых измерений вводятся прокси-переменные, аппроксимирующие ненаблюдаемые факторы. На качественном уровне оценивается риск мультиколлинеарности, предусматривается включение фиктивных переменных для учёта качественных факторов, а также постулируются возможные парные и многофакторные взаимодействия между регрессорами. До проведения эмпирического анализа задаются нелинейные преобразования предикторов и формулируются теоретические ограничения на знаки оценок коэффициентов, вытекающие из экономической, физической или иной предметной логики.

3. Подсистема «Сбор статистических данных» для получения, очистки, обработки пропусков и выбросов, разделения выборки и документирования границ применимости. В этой подсистеме осуществляется выбор источников информации, обоснование необходимого объёма выборки с учётом требуемой мощности критериев, а для временных рядов дополнительно определяются частота регистрации и протяжённость ретроспективного периода. Проверяется репрезентативность полученной совокупности относительно генеральной выборки, после чего выполняется первичная очистка дан-

ных. Проводится выявление и маркировка аномальных наблюдений с использованием статистических критериев и визуальных методов. Выполняется диагностика мультиколлинеарности на основе корреляционной матрицы предикторов и рассчитываются показатели дисперсии независимых переменных. Завершающей операцией выступает документирование границ применимости модели — диапазонов изменения факторов, за пределами которых использование регрессионной зависимости становится некорректным.

4. Подсистема «Спецификация модели». В этой подсистеме осуществляется выбор класса регрессионного уравнения, определяемый типом целевой переменной, что предопределяет применение, соответственно, линейной, логистической, порядковой или пуассоновской регрессии. Устанавливается характер связи между откликом и независимыми переменными. При этом принимаются решения о включении членов взаимодействия между предикторами и о введении нелинейных преобразований регрессоров для адекватного описания криволинейных зависимостей. Дополнительно задается структура случайной составляющей. Завершается данный этап записью формальной математической спецификации в виде регрессионного уравнения, содержащего все структурные элементы — параметры, переменные, преобразования и вероятностные допущения.

5. Подсистема «Идентификация модели». В этой подсистеме осуществляется выбор метода оценивания неизвестных параметров, который определяется типом целевой переменной, видом спецификации и предположениями о распределении ошибок. После обоснования выбранного критерия оптимизации выполняются численные расчёты точечных оценок коэффициентов регрессии, обеспечивающие минимизацию целевой функции или максимизацию функции правдоподобия на основе имеющейся обучающей выборки.

6. Подсистема «Верификация модели» для проведения комплексной статистической диагностики оценённого регрессионного уравнения. В первую очередь проверяется содержательная согласованность полученных результатов путём сопоставления фактических знаков оценок коэффициентов с теоретическими ожиданиями, затем рассчитываются стандартные ошибки коэффициентов, на основе которых проводится проверка их значимости с использованием t-критерия Стьюдента и строятся соответствующие доверительные интервалы. Оценивается общее качество аппроксимации посредством соответствующих метрик, а значимость модели в целом проверяется с помощью F-критерия Фишера. В целях обеспечения надёжности выводов исследуется чувствительность коэффициентов к удалению отдельных наблюдений, рассчитываются показатели VIF для уточнённой диагностики мультиколлинеарности, а также выполняется анализ устойчивости оценок на

подвыборках. Особое внимание уделяется проверке стандартных предположений относительно случайной составляющей.

7. Подсистема «Интерпретация, принятие решения и подготовка отчёта». В этой подсистеме выполняется содержательная интерпретация оценённых коэффициентов, предельных эффектов и прогнозных значений в терминах предметной области, что позволяет оценить практическую значимость модели и её соответствие изначально поставленной задаче. Проводится сопоставление полученных результатов с теоретическими гипотезами, бизнес-ограничениями и критериями допустимой ошибки. При обнаружении неудовлетворительных характеристик принимается решение о возврате к более ранним этапам для доработки модели. В случае положительного решения о достижении целевых показателей формируется итоговый отчёт для заказчика, содержащий формализованное описание спецификации, результаты оценки и верификации, доверительные интервалы, границы применимости, а также обоснованные рекомендации по использованию модели, что завершает полный цикл регрессионного моделирования.

Каждая подсистема выполняет уникальную функцию и не может быть исключена без деградации системы. Подсистемы связаны не произвольно, а в определенной последовательности, при которой результирующий выход каждого предшествующего этапа служит входным аргументом для последующего, формируя тем самым конвейерную структуру обработки информации. Конфигурация системы допускает обратные итерационные связи между подсистемами.

На схеме (рис. 1) показано 5 блоков *внешней среды*, окружающих систему.

1. *Фундаментальные законы*. Это объективная реальность, в которой существует система. Сюда входят законы термодинамики, демографические тренды, культурные нормы, исторический контекст, экологические ограничения и т.д. Этот блок задает фундаментальную форму связей между переменными. В физике это могут быть степенные законы (квадрат расстояния), в биологии — логистические кривые, в социологии — пороговые эффекты. Если система выберет линейную спецификацию там, где в среде действует нелинейный закон, модель будет принципиально неадекватна.

2. *Информационно-измерительная среда*. Это всё, что связано с тем, откуда система берет данные для этапа 3. Сюда входят датчики со своими погрешностями, периодичность измерений, глубина исторического архива, наличие пропусков, способы агрегации данных, методы опроса, протоколы экспериментов и т.д. Если данные собираются с большим шумом, например, спутниковые данные в условиях облачности, то коэффициенты регрессии будут иметь огромные стандартные ошибки, и интерпретация ста-

нет ненадежной. Если данные измеряются редко, например, раз в год, то модель не сможет уловить быстрые процессы.

3. *Человеческий фактор*. Он вносит в систему субъективные цели, предпочтения, ограничения и когнитивные искажения. Сюда входит квалификация команды, теоретические пристрастия экспертов, административные сроки, бюджет, карьерные интересы заказчика, психологическая инерция («мы всегда так делали») и т.д. Если заказчик нетерпелив, система будет вынуждена быстро строить неадекватную модель. Если эксперты в команде конфликтуют, например, первый считает важным фактор A, а второй — фактор B, то это может парализовать этап 2. Кроме того, человек может сознательно или бессознательно «подгонять» модель под желаемый результат, и тогда система производит ложную интерпретацию.

4. *Технологическая среда*. Это материально-цифровая база, которая позволяет системе реализовать этапы 4–6. Сюда входят вычислительные мощности, программное обеспечение, алгоритмические библиотеки, инфраструктура и т.д. Всё это накладывает технические ограничения на сложность модели. Если в данных миллиарды строк, а у исполнителей в распоряжении обычный ноутбук с 8 ГБ ОЗУ, то оценить регрессию с тысячами фиктивных переменных не получится. Внешняя среда вынудит систему использовать упрощенные алгоритмы. Кроме того, технологии могут устаревать. Если в распоряжении исследователей старые программы, то это не позволит реализовать новые методы.

5. *Временная динамика*. Это среда, которая делает процесс построения модели принципиально незавершенным. Даже идеально построенная модель сегодня перестает быть адекватной завтра, потому что меняется сам объект или условия его функционирования. После того как модель построена (этап 7), внешняя временная среда начинает ее «разрушать». Это заставляет систему работать не однократно, а циклически: вернуться к этапу 3 (собрать свежие данные) и переоценить модель. Например, регрессия, прогнозирующая успеваемость школьников по результатам тестов 2010 г., к 2026 г. становится бесполезна, потому что изменилась образовательная среда и старые факторы перестали быть значимыми.

В системном анализе входы — это то, что система потребляет из внешней среды для своей работы, а выходы — то, что она возвращает обратно в среду.

Система на схеме (рис. 1) имеет 4 *входа*.

1. *Запрос на решение проблемы*. Это исходный импульс, исходящий от заказчика, который запускает весь процесс. Это «боль», «вопрос» или «необходимость», существующая во внешней среде до начала любого моделирования. Запрос представляет собой текст на естественном языке: «Почему это происходит?», «Как будет дальше?», «Что будет, если я изменю параметр?».

2. *Эмпирические данные*. Это информация о прошлом состоянии объекта исследования. Данные представляют собой массивы чисел, кодируемые текстовые категории, временные метки. Зачастую такие данные содержат шум, пропуски и выбросы. Данные поступают из информационно-измерительного и технологического блока внешней среды.

3. *Гипотезы*. Это априорные знания, которые не содержатся в данных, и теоретические ограничения, которые позволят сделать модель осмысленной. Сюда поступают ожидаемые знаки коэффициентов, известные нелинейности, список важных факторов, запрет на некоторые переменные и знания о причинно-следственных структурах. Информация поступает из блока «Фундаментальные законы» и «Человеческий фактор» (опыт экспертов).

4. *Ресурсные ограничения*. Сюда относятся ограничения на ресурсы времени, вычислений и денег. Без этих ограничений система теоретически могла бы строить бесконечно сложную модель вечно, но реальность не позволяет. Сюда также поступает допустимая сложность модели: простая — для быстрого внедрения, сложная — для точного прогноза. Информация поступает из блока «Технологическая среда» (вычислительная мощность) и «Человеческий фактор» (смета проекта).

Система на схеме (рис. 1) имеет 4 *выхода*.

1. *Уравнение регрессии и статистики*. Это главный математический продукт, который производит система в техническом смысле. Информация направляется в технологическую среду (внедряется в приложение) и в руки заказчика в виде отчёта.

2. *Результаты интерпретации*. Здесь математические формулы превращаются в понятные слова. Именно этот выход делает модель полезной для принятия решений. Конкретно даётся ответ на исходный вопрос «фактор x значимо влияет на y или нет?», констатируется величина эффекта «при изменении x на единицу y меняется на столько-то», проводится ранжирование факторов по степени влияния на y и делается заключение о подтверждении или опровержении теоретической гипотезы. Информация направляется, например, в научную статью или для принятия управленческого решения. Также эта информация может влиять на «Фундаментальные законы» внешней среды: если гипотеза подтверждена, то она укрепляет теорию, а если опровергнута, то теорию пересматривают.

3. *Прогнозы*. В этой части выхода система экстраполирует найденные связи в будущее или в альтернативные условия. Конкретно выдаются точечные прогнозы, интервальные прогнозы, а также результаты сценарного анализа: «что будет, если x_1 вырастет на 10 %, а x_2 останется неизменным?». Информация направляется в блоки «Информационно-измерительная среда» и «Временная динамика» для сравнения в будущем с реальными данными и калибровки системы.

3. *Рекомендации по улучшению*. Здесь система возвращает во внешнюю среду критику самой среды — указание на то, что мешало строить лучшую модель. Например, если в процессе построения модели установлено, что «модель неадекватна, потому что в данных нет x_3», то может быть сформулирована рекомендация «установить новый датчик или добавить вопрос в опросник». Или, например, установлено, что «модель переобучена». Тогда рекомендация «собрать больше наблюдений». Информация направляется в информационно-измерительную среду, меняя протоколы сбора данных, в технологическую среду, улучшая инфраструктуру, а также в блок «Человеческий фактор», повышая компетенции.

На рис. 1 черной стрелкой показано замыкание контура *обратной связи*, реализующее передачу воздействия с выходов на входы. Причем, можно выделить 4 петли обратной связи.

1. Поведенческая обратная связь, направленная от выходов ко входу 1. Её суть в том, что применение модели изменяет поведение объекта или цели заказчика. Проблема, которая была на входе, перестает существовать или трансформируется в новую. Система вынуждена запускаться заново с новым запросом.

2. Инструментальная обратная связь, направленная от выходов ко входу 2. Результаты модели показывают, каких данных не хватает или какие данные собираются с ошибками. Система выдает рекомендации по изменению протоколов измерения, и в следующий цикл поступают уже более релевантные данные.

3. Когнитивная обратная связь, направленная от выходов ко входу 3. Результаты модели изменяют теоретические представления заказчика о предметной области. То, что раньше было гипотезой, становится подтвержденным фактом или опровергнутым заблуждением. И в следующий цикл заказчик входит уже с новыми априорными знаниями.

4. Технологическая обратная связь, направленная от выходов ко входу 4. Опыт применения модели показывает, какие вычислительные ресурсы и время реально нужны для достижения приемлемого качества. В следующий цикл система запрашивает или получает больше ресурсов — или, наоборот, учится работать с меньшими.

Также на рис. 1 указаны 3 участника процесса построения регрессионной модели.

1. Эксперт (внутри системы). Это «переводчик» между математическими конструкциями и реальным миром. Он обеспечивает содержательную валидность модели.

2. Аналитик (внутри системы). Он обеспечивает статистическую корректность, надежность и воспроизводимость результатов. Владеет инструментарием регрессионного анализа.

3. Заказчик (во внешней среде). Он задаёт требования и оценивает выходной продукт, но не участвует во внутренних процедурах построения модели.

На основании проведённого анализа можно утверждать, что процесс построения регрессионной модели обладает всеми необходимыми признаками системы в терминах системного подхода. Во-первых, он обладает целенаправленностью, поскольку ориентирован на достижение цели, заданной заказчиком. Во-вторых, он структурирован и иерархичен: распадается на взаимосвязанные подсистемы (этапы), которые, в свою очередь, могут делиться на отдельные операции. В-третьих, процесс является открытой системой, обменивающейся с внешней средой информацией и ресурсами. В-четвертых, зафиксировано наличие чётких входов и выходов. В-пятых, выявлен контур обратной связи.

2. Выделение элементов системы, влияющих на интерпретируемость результирующей модели

Допустим, что после 6-го этапа «Верификация модели» получена регрессионная модель с малыми значениями ошибок аппроксимации и со значимыми на заданном уровне коэффициентами. Тем не менее, использование полученной регрессионной зависимости в реальной практике способно приводить к негативным, вплоть до катастрофических, последствиям. Для иллюстрации данного тезиса рассмотрим два характерных примера.

Пример 1. Клиническая эндокринология.

Рассмотрим задачу персонализированного подбора суточной дозы инсулина у пациентов с сахарным диабетом. Допустим, что на основе выборочных клинических данных построена модель множественной линейной регрессии следующего вида:

$$y = 10 + 0{,}5x_1 + 2{,}2x_2 + 0{,}05x_3 - x_4 + 0{,}8x_5, \qquad (1)$$

где y — суточная доза инсулина (ЕД/сут); x_1 — возраст пациента; x_2 — уровень глюкозы натощак (ммоль/л); x_3 — концентрация гликированного гемоглобина HbA1c; x_4 — уровень физической активности; x_5 — индекс массы тела.

Коэффициент при уровне глюкозы в модели (1) означает предписываемое увеличение суточной дозы инсулина на 2,2 ЕД при повышении x_2 на 1 ммоль/л. Предположим, что величина 2,2 существенно превышает диапазон, например, 0,5 — 1,0 ЕД/сут на 1 ммоль/л, рекомендованный действующими клиническими протоколами, поэтому в клиническом контексте является потенциально опасной. Например, реализация модели (1) в цифровом приложении-ассистенте может привести к тому, что алгоритм для измеренного уровня глюкозы сформирует рекомендацию по уве-

личению дозы инсулина, превышающую физиологически безопасный коридор. Эта рекомендация приведёт к риску гипогликемии, сопровождающейся утратой сознания, что квалифицируется как угрожающее жизни человека состояние.

Причинами столь высокой и клинически неадекватной оценки коэффициента могут быть: 1) мультиколлинеарность, например, высокая корреляция между глюкозой и HbA1c; 2) пропущенная переменная, например, инсулинорезистентность; 3) влияние выбросов, например, наличие в выборке пациентов с высокой глюкозой и очень высокими дозами инсулина, которые смещают оценку в сторону завышения.

Пример 2. Промышленное машиностроение.

Рассмотрим регрессионную модель, специфицированную для прогнозирования интенсивности износа деталей металлообрабатывающего станка в зависимости от условий эксплуатационной среды. Пусть оцененное уравнение имеет следующий вид:

$$y = 10 - 3{,}5x_1 + 1{,}2x_2 + 0{,}1x_3, \qquad (2)$$

где y — износ (в относительных единицах); x_1 — температура воздуха в цеховом помещении (°C); x_2 — давление в гидравлической системе; x_3 — среднесуточная относительная влажность воздуха.

При внедрении модели (2) в систему поддержки принятия решений на предприятии может возникнуть следующая опасная ситуация. Инженер по эксплуатации, руководствуясь отрицательным коэффициентом при температуре x_1, может принять ошибочное решение о повышении температурного режима в цеху с целью якобы уменьшить износ и продлить ресурс деталей. В действительности повышение температуры не приводит к ожидаемому снижению износа, а наоборот, ускоряет деградацию материалов за счёт снижения вязкости смазочных материалов и активации окислительных реакций. В конечном счёте, данная управленческая интервенция обернётся увеличением фактического износа сверх расчётных величин, а также внеплановыми остановками и, как следствие, экономическими потерями.

Данная ситуация может быть обусловлена следующими причинами: 1) мультиколлинеарность, например, сильная отрицательная корреляционная связь между температурой и влажностью; 2) неучёт лаговых переменных, например, неучёт инерционности тепловых процессов.

Таким образом, интерпретируемость выделяется как системное свойство регрессионных моделей, имеющее критическое значение для их практической применимости. Модели (1) и (2) данным свойством не обладают, что исключает возможность формирования доверия к их результатам и, следовательно, делает их непри-

годными для использования в целях принятия управленческих решений. Под интерпретируемостью в данном контексте понимается наличие обоснованного доверия к модели, в первую очередь — к статистической и содержательной значимости её коэффициентов. Согласно схеме на рис. 1, свойство интерпретируемости формируется на выходе подсистемы 6 «Верификация модели».

Подчеркнём, что интерпретируемость носит эмерджентный характер. Она не сводится к сумме свойств отдельных этапов построения модели и не может быть гарантирована качественным выполнением каждого из них в отдельности. Данное свойство возникает исключительно на системном уровне — как интегральный эффект целостного процесса. Иными словами, интерпретируемость проявляется как неожидаемое системное качество, которое либо порождается, либо не возникает вовсе, даже при формально корректной реализации всех предшествующих подсистем.

С методологической точки зрения, интерпретируемость является не только желаемым, но и управляемым атрибутом модели, что требует целенаправленного контроля над её формированием. В рамках системного подхода это обусловливает необходимость декомпозиции каждого из этапов 1–6 на элементарные составляющие. Детализация подсистем позволит перейти от констатации наличия или отсутствия интерпретируемости к выявлению конкретных причин, её определяющих.

В табл. представлены результаты декомпозиции этапов 1–6 на элементарные составляющие — операции. При этом последние не являются предельно элементарными и в принципе допускают дальнейшее дробление. Однако для целей диагностики причин утраты или отсутствия интерпретируемости предложенного уровня детализации представляется достаточным. В табл. операции, выделенные зеленым цветом, контролируются экспертом, оранжевым цветом — аналитиком, белым цветом — обоими. Как видно, на этапах 1–2 преобладает деятельность эксперта, а на этапах 3–6 — аналитика.

Анализ представленных в таблице 1 операций позволил выделить те из них, которые, по мнению автора, наиболее сильно влияют на интерпретируемость регрессионных моделей. Кратко перечислим их далее.

1.4. Определение допустимой ошибки — ответ на вопрос: «Насколько высоким должно быть качество аппроксимации модели?».

1.5. Выбор приоритета: точность, интерпретируемость или простота модели — ответ на вопрос: «Если придется выбирать, то чем мы пожертвуем в первую очередь?». Отметим, что этапы 1.3 и 1.5 не одно и то же. Например, на этапе 1.3 может быть выбрана задача прогнозирования, а на этапе 1.5 приоритет — интерпретируемость (жертвуем точностью). В таком случае предпочтение стоит отдать линейным регрессиям.

1.7. Выявление ограничений по вычислительным ресурсам и времени — это ответ на вопрос: «Сколько у нас есть мощностей и времени для расчётов?». Здесь определяется, можно ли использовать «тяжёлые» методы, например, для построения сложных нелинейных моделей, или придётся ограничиться быстрыми. Если ресурсов мало, то выбор падает на простые методы.

1.8. Определение требований к объяснимости — ответ на вопрос: «Должен ли каждый коэффициент иметь прямой экономический/физический смысл или нет?». Если да, то коэффициенты должны трактоваться как «при изменении x на 1 единицу y меняется на …» (содержательная интерпретация). Этот пункт запрещает сильно коррелированные переменные, а также использование полиномов высоких степеней. Если нет, то можно использовать регуляризацию, нелинейные преобразования, взаимодействия, даже если они бессмысленны сами по себе.

1.9. Установление порога статистической значимости — ответ на вопрос: «С какой вероятностью мы готовы ошибиться, признав эффект значимым, когда его на самом деле нет?». Задается уровень α (обычно 0,05, 0,01 или 0,1) — допустимая вероятность ошибки первого рода (ложноположительного вывода). Чрезмерно жесткий уровень значимости ($\alpha = 0{,}001$) повышает риск ошибки второго рода — исключения значимых, но слабо выраженных факторов. Напротив, излишне мягкий порог ($\alpha = 0{,}1$) ведет к включению в модель случайных шумов, интерпретируемых как статистически значимые зависимости (ошибка первого рода).

2.1. Составление теоретического списка всех потенциальных предикторов — ответ на вопрос: «Какие факторы гипотетически могут влиять на целевой показатель, исходя из теории и здравого смысла?». Составляется максимально широкий перечень переменных, которые, по мнению эксперта, способны оказывать влияние на зависимую переменную. Полнота этого списка напрямую влияет на корректность последующей интерпретации, поскольку пропуск теоретически значимого фактора на начальном этапе делает невозможным его включение в модель и приводит к смещению оценок влияния оставшихся переменных.

2.6. Проверка на мультиколлинеарность на уровне здравого смысла — ответ на вопрос: «Нет ли среди факторов таких, которые логически измеряют одну и ту же сущность, даже если формально это разные переменные?». Если есть — включать их вместе нельзя: коэффициенты станут бессмысленными, модель не поймёт, чему приписать эффект, а заказчик получит противоречивую интерпретацию. Лучше заранее оставить один фактор, который теоретически важнее.

2.9. Введение нелинейных преобразований объясняющих переменных до анализа данных — ответ на вопрос: «Может ли влияние фактора на результат быть нелинейным, и как это учесть

ISSN 2713-1734

Операции подсистем 1–6

№	Операции										
1	1.1. Формирование портрета заказчика модели	1.2. Выбор целевого показателя (объясняемой переменной)	1.3. Выбор типа решаемой задачи	1.4. Определение допустимой ошибки	1.5. Выбор приоритета: точность, интерпретируемость или простота модели	1.6. Определение структуры данных	1.7. Выявление ограничений по вычислительным ресурсам и времени	1.8. Определение требований к объяснимости	1.9. Установление порога статистической значимости	1.10. Определение необходимости доверительных интервалов для коэффициентов и прогнозов	1.11. Согласование формата отчёта
2	2.1. Составление теоретического списка всех потенциальных предикторов (объясняющих переменных)	2.2. Классификация объясняющих переменных по шкалам	2.3. Выявление лаговых переменных (если временной ряд)	2.4. Определение инструментальных переменных (для устранения эндогенности)	2.5. Выявление прокси-переменных	2.6. Проверка на мультиколлинеарность на уровне здравого смысла	2.7. Учёт фиктивных переменных	2.8. Определение возможных взаимодействий	2.9. Введение нелинейных преобразований объясняющих переменных до анализа данных	2.10. Формирование теоретических ограничений на знаки коэффициентов при переменных	
3	3.1. Определение источника данных	3.2. Определение объёма выборки	3.3. Определение частоты и длины периода сбора (для временных рядов)	3.4. Проверка репрезентативности выборки	3.5. Первичная очистка	3.6. Обработка пропусков	3.7. Выявление и маркировка выбросов	3.8. Разбивка на обучающую и тестовую выборки	3.9. Диагностика мультиколлинеарности по корреляционной матрице	3.10. Оценка дисперсии независимых переменных	3.11. Документация границ применимости модели

Окончание табл.

4	4.1. Выбор класса регрессии исходя из типа целевой переменной	4.2. Определение линейной/нелинейной связи	4.3. Решение о включении взаимодействий	4.4. Решение о включении нелинейных преобразований объясняющих переменных	4.5. Выбор формы ошибки	4.6. Определение распределения ошибок	4.7. Запись формальной математической спецификации (уравнение) до оценки параметров				
5	5.1. Выбор метода оценки параметров модели	5.2. Расчёт точечных оценок коэффициентов									
6	6.1. Проверка согласованности теоретических и фактических знаков оценок	6.2. Расчёт стандартных ошибок коэффициентов	6.3. Проверка значимости коэффициентов по t-критерию Стьюдента	6.4. Расчёт доверительных интервалов для коэффициентов	6.5. Расчёт меры качества аппроксимации	6.6. Проверка значимости модели в целом по F-критерию Фишера	6.7. Оценка чувствительности коэффициентов к удалению наблюдений	6.8. Расчёт VIF для диагностики мультиколлинеарности	6.9. Графический анализ остатков	6.10. Тесты на гомоскедастичность	6.11. Тесты на автокорреляцию остатков
	6.12. Тест на нормальность остатков	6.13. Прогнозирование на тестовой выборке	6.14. Проверка стабильности коэффициентов	6.15. Анализ остаточной дисперсии по группам							

в структуре модели до анализа данных?». На основе предметной области или здравого смысла задаются функциональные формы: логарифмы, полиномы второй/третьей степени (для U-образных зависимостей), обратные преобразования и т.д. Включение преобразований повышает сложность интерпретации.

2.10. Формирование теоретических ограничений на знаки коэффициентов при переменных — ответ на вопрос: «Какие знаки коэффициентов при факторах должны получиться, чтобы модель не противоречила теории?». Здесь формулируются ожидаемые направления влияния объясняющих переменных на целевой показатель, вытекающие из логики предметной области. Эти ограничения служат содержательным критерием валидности модели: оценённый коэффициент должен не только быть статистически значимым, но и иметь знак, согласующийся с априорными ожиданиями. Если знак коэффициента противоречит теории, это свидетельствует о наличии систематического смещения и делает содержательную интерпретацию данного фактора некорректной. Таким образом, априорные ограничения на знаки являются не техническим допущением, а инструментом проверки того, что модель отражает реальные причинно-следственные связи, а не случайные корреляции.

3.4. Проверка репрезентативности выборки — ответ на вопрос: «Насколько наша выборка похожа на ту совокупность, для которой мы будем делать выводы?». Проверка включает анализ возможных источников систематического смещения: географических, временных, канальных, демографических и др. Репрезентативность напрямую определяет границы интерпретации: если выборка смещена (например, провели опрос только иркутян), коэффициенты модели отражают закономерности не для всей совокупности, а лишь для её части. В этом случае интерпретация должна сопровождаться явным ограничением области применения результатов.

3.5. Первичная очистка — ответ на вопрос: «Есть ли в данных явные технические ошибки, которые искажают реальную картину?». Здесь проводится минимальная предобработка данных: удаляются полностью повторяющиеся записи (дубликаты), исправляются или исключаются наблюдения с явными ошибками ввода (например, возраст = 200 лет, отрицательная цена и т.д.). Роль этого этапа в интерпретации состоит в том, чтобы гарантировать, что оцениваемые коэффициенты отражают реальные закономерности, а не являются результатом технических сбоев или опечаток. Пропуск этого этапа делает все последующие выводы потенциально ошибочными.

3.7. Выявление и маркировка выбросов — ответ на вопрос: «Какие наблюдения резко выбиваются из общего паттерна данных?». Выбросы ищут, например, с помощью визуализации («ящики с усами», точечные графики). Каждый выброс не удаляют автоматически, а маркируют для дальнейшего решения. Вы-

бросы искажают коэффициенты регрессии (особенно при МНК), что негативно сказывается на интерпретируемости. Маркировка позволяет в итоговом отчёте отдельно описать, например, что «коэффициенты получены без учёта трёх аномальных регионов», что делает интерпретацию честной и воспроизводимой.

3.9. Диагностика мультиколлинеарности по корреляционной матрице — ответ на вопрос: «Есть ли между предикторами парные корреляции, которые могут указывать на дублирование информации и потенциальные проблемы с разделением эффектов?». Высокая парная корреляция служит ранним индикатором потенциальной мультиколлинеарности, при которой коэффициенты регрессии становятся нестабильными, а их интерпретация как мера независимого влияния факторов — некорректной.

4.1. Выбор класса регрессии исходя из типа целевой переменной — ответ на вопрос: «Какова природа целевой переменной, и какая функция связи лучше всего описывает её зависимость от факторов?». Исходя из типа целевой переменной y, выбирается класс регрессионной модели: линейная (тип y — непрерывный), логистическая (тип y — бинарный), пуассоновская (тип y — счётчик), мультиномиальная (тип y — многоклассовый). Данный выбор фундаментальным образом задаёт способ интерпретации коэффициентов. Неправильно выбранный тип модели делает интерпретацию либо технически некорректной, либо содержательно бессмысленной, даже если статистические критерии формально выполнены.

4.2. Определение линейной/нелинейной связи — ответ на вопрос: «Является ли связь между факторами и целевой переменной линейной?». Если нет, то применяются нелинейные преобразования (полиномы, сплайны, преобразования Бокса-Кокса), чтобы модель лучше отражала реальную зависимость. Интерпретация в таком случае сложнее.

4.4. Решение о включении нелинейных преобразований объясняющих переменных принимается исходя из пункта 2.9.

5.1. Выбор метода оценки параметров модели — ответ на вопрос: «По какому принципу мы будем подбирать коэффициенты модели, чтобы они были максимально надёжными и соответствовали природе данных?». Основные методы: МНК, ММП, МНМ, метод моментов и др. Метод определяет, можно ли верить интерпретации коэффициентов как реальных эффектов.

5.2. Расчёт точечных оценок коэффициентов — ответ на вопрос: «Какое числовое значение принимает каждый коэффициент в нашей спецификации?». Эти оценки являются количественным ядром последующей интерпретации.

6.1. Проверка согласованности теоретических и фактических знаков оценок — ответ на вопрос: «Не противоречат ли полученные знаки коэффициентов тому, что предсказывает теория?». Если

знак совпадает — модель ведёт себя содержательно логично, отражая реальные причинно-следственные связи; если нет — это сигнал о возможных проблемах (эндогенность, пропущенные переменные, ошибки спецификации). На проверку влияет пункт 2.10.

6.3. Проверка значимости коэффициентов по t-критерию Стьюдента — ответ на вопрос: «Можно ли с достаточной уверенностью утверждать, что коэффициент действительно отличается от нуля, или полученное значение — это просто случайность выборки?». Значимость коэффициента говорит только о его неслучайной природе, но не эквивалентна важности. На больших выборках значимыми становятся даже микроскопические эффекты, которые на практике могут быть бесполезны.

6.6. Проверка значимости модели в целом по F-критерию Фишера — ответ на вопрос: «Объясняет ли модель в целом целевую переменную лучше, чем простая константа (среднее значение)?». Статистическая значимость F-критерия является необходимым, хотя и недостаточным, условием для последующей интерпретации индивидуальных коэффициентов: если модель в целом незначима, содержательные выводы по отдельным факторам не могут считаться обоснованными.

6.7. Оценка чувствительности коэффициентов к удалению наблюдений — ответ на вопрос: «Насколько сильно меняются коэффициенты, если убрать из выборки отдельные наблюдения?». Если коэффициенты сильно меняются при удалении одного-двух наблюдений — оценки неустойчивы, и интерпретировать их как надёжные нельзя.

6.8. Расчёт VIF для диагностики мультиколлинеарности — ответ на вопрос: «Не слишком ли сильно факторы в модели связаны друг с другом, чтобы можно было доверять их раздельным коэффициентам?». Значения VIF > 10 (в ряде случаев > 5) указывают на высокую коррелированность фактора с другими регрессорами, что приводит к нестабильности оценок коэффициентов и завышению их стандартных ошибок. В таких условиях интерпретация коэффициентов как меры раздельного влияния отдельных факторов становится некорректной, поскольку модель не способна идентифицировать их независимые эффекты.

Заключение

В ходе проведённого исследования обосновано, что процесс построения регрессионной модели с позиций системного анализа обладает всеми необходимыми признаками системы. Кроме того, в работе идентифицированы ключевые факторы, определяющие интерпретируемость регрессионной модели. Представленное системное описание и выявленные факторы интерпретируемости создают методологическую основу для разработки процедур управления процессом построения регрессионных моделей.

Список использованной литературы

1. Molnar C. Interpretable machine learning / C. Molnar. — Lulu.Com, 2020. — 251 p.

2. Interpretable machine learning in physics: a review / S.J. Wetzel, S. Ha, R. Iten [et al] // ResearchGate. — 2025. — URL: https://www.researchgate.net/publication/390354224_Interpretable_Machine_Learning_in_Physics_A_Review. — DOI 10.48550/arXiv.2503.23616.

3. Challenging the performance-interpretability trade-off: An evaluation of interpretable machine learning models / S. Kruschel, N. Hambauer, S. Weinzierl [et al] — DOI 10.1007/s12599-024-00922-2 // Business & Information Systems Engineering. — 2026. — Vol. 68, no. 1. — P. 159–183.

4. Interpretable machine learning for statistical modeling: Bridging classical and modern approaches / R. Khan, A.M. Shah, Bareera [et al]. — DOI 10.5281/zenodo.16737426 // International Journal of Social Sciences Bulletin. — 2025. — Vol. 3, no. 8. — P. 43–50.

5. Hoang N.D. From data to insights: Modeling urban land surface temperature using geospatial analysis and interpretable machine learning / N.D. Hoang, V.D. Tran, T.C. Huynh. — DOI 10.3390/s25041169 // Sensors. — 2025. — Vol. 25, no. 4. — P. 1169.

6. Leveraging interpretable machine learning in intensive care / L. Bohlen, J. Rosenberger, P. Zschech, M. Kraus. — DOI 10.1007/s10479-024-06226-8 // Annals of Operations Research. — 2025. — Vol. 347, no. 2. — P. 1093–1132.

7. Durap A. Interpretable machine learning for coastal wind prediction: Integrating SHAP analysis and seasonal trends / A. Durap. — DOI 10.1007/s11852-025-01108-y // Journal of Coastal Conservation. — 2025. — Vol. 29, no. 3. — P. 24.

8. Makke N. Interpretable scientific discovery with symbolic regression: a review / N. Makke, S. Chawla. — DOI 10.1007/s10462-023-10622-0 // Artificial Intelligence Review. — 2024. — Vol. 57, no. 1. — P. 2.

9. Salih A.M. Are linear regression models white box and interpretable? / A.M. Salih, Y. Wang. — DOI 10.22541/au.172122905.57636646/v1 // ResearchGate. — 2024. — URL: https://www.researchgate.net/publication/382331613_Are_Linear_ Regression_Models_White_Box_and_Interpretable.

10. Горбач А.Н. Покупательское поведение: анализ спонтанных последовательностей и регрессионных моделей в маркетинговых исследованиях / А.Н. Горбач, Н.А. Цейтлин. — Киев, 2011. — 220 с.

11. Базилевский М.П. Построение вполне интерпретируемых линейных регрессионных моделей с помощью метода последовательного повышения абсолютных вкладов переменных в общую детерминацию / М.П. Базилевский — DOI 10.17308/sait/1995-5499/2022/2/5-16. — EDN CNDSSW // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. — 2022. — № 2. — С. 5–16.

12. Базилевский М.П. Программа построения вполне интерпретируемых элементарных и неэлементарных квазилинейных регрессионных моделей / М.П. Базилевский. — DOI 10.15514/ISPRAS-2023-35(4)-7. — EDN KTOSCW // Труды Института системного программирования РАН. — 2023. — Т. 35, № 4. — С. 129–144.

13. Базилевский М.П. О разрешимости оптимизационной задачи построения вполне интерпретируемых линейных регрессий / М.П. Базилевский. — DOI 10.26907/2541-7746.2025.4.627-640. — EDN PXUNAW // Ученые записки Казанского университета. Серия: Физико-математические науки. — 2025. — Т. 167, № 4. — С. 627–640.

14. Базилевский М.П. Метод построения вполне интерпретируемых линейных регрессий со статистически значимыми по критерию Стьюдента оценками и незначимыми коэффициентами интеркорреляции / М.П. Базилевский. — DOI 10.26089/NumMet.v26r435. — EDN WKVQKH // Вычислительные методы и программирование. — 2025. — Т. 26, № 4. — С. 534–547.

15. Оптнер С.Л. Системный анализ для решения деловых и промышленных проблем / С.Л. Оптнер. — Москва : Сов. радио, 1969. — 216 с.

16. Янг С. Системное управление организацией / С. Янг. — Москва : Сов. радио, 1972. — 456 с.

17. Федоренко Н.П. О методах социально-экономического прогнозирования / Н.П. Федоренко // Методология прогнозирования экономического развития СССР. — Москва, 1971. — С. 9–21.

18. Черняк Ю.И. Системный анализ в управлении экономикой / Ю.И. Черняк. — Москва : Экономика, 1975. — 191 с.

19. Перегудов Ф.И. Основы системного проектирования АСУ организационными комплексами / Ф.И. Перегудов. — Томск : Изд-во Том. ун-та, 1984. — 177 с.

20. Айвазян С.А. Прикладная статистика и основы эконометрики / С.А. Айвазян, В.С. Мхитарян. — Москва : ЮНИТИ, 1998. — 1003 с.

References

1. Molnar C. *Interpretable Machine Learning*. Lulu.Com, 2020. 251 p.

2. Wetzel S.J., Ha S., Iten R., Klopotek M., Liu Z. Interpretable Machine Learning in Physics: a Review. *ResearchGate*, 2025. Available at: https://www.researchgate.net/publication/390354224_Interpretable_Machine_Learning_in_Physics_A_Review. DOI: 10.48550/arXiv.2503.23616.

3. Kruschel S., Hambauer N., Weinzierl S., Zilker S., Kraus M., Zschech P. Challenging the Performance-Interpretability Trade-off: An Evaluation of Interpretable Machine Learning Models. *Business & Information Systems Engineering*, 2026, vol. 68, no. 1, pp. 159–183. DOI: 10.1007/s12599-024-00922-2.

4. Khan R., Shah A.M., Bareera, Ijaz A., Sumeer A. Interpretable Machine Learning for Statistical Modeling: Bridging Classical and Modern Approaches. *International Journal of Social Sciences Bulletin*, 2025, vol. 3, no. 8, pp. 43–50. DOI: 10.5281/zenodo.16737426.

5. Hoang N.D., Tran V.D., Huynh T.C. From Data to Insights: Modeling Urban Land Surface Temperature Using Geospatial Analysis and Interpretable Machine Learning. *Sensors,* 2025, vol. 25, no. 4, pp. 1169. DOI: 10.3390/s25041169.

6. Bohlen L., Rosenberger J., Zschech P., Kraus M. Leveraging Interpretable Machine Learning in Intensive Care. *Annals of Operations Research,* 2025, vol. 347, no. 2, pp. 1093–1132. DOI: 10.1007/s10479-024-06226-8.

7. Durap A. Interpretable Machine Learning for Coastal Wind Prediction: Integrating SHAP Analysis and Seasonal Trends. *Journal of Coastal Conservation*, 2025, vol. 29, no. 3, pp. 24. DOI: 10.1007/s11852-025-01108-y.

8. Makke N., Chawla S. Interpretable Scientific Discovery with Symbolic Regression: a Review. *Artificial Intelligence Review*, 2024, vol. 57, no. 1, pp. 2. DOI: 10.1007/s10462-023-10622-0.

9. Salih A.M., Wang Y. Are Linear Regression Models White Box and Interpretable? *ResearchGate*. 2024. Available at: https://www.researchgate.net/publication/382331613_Are_Linear_Regression_Models_White_Box_and_Interpretable. DOI 10.22541/au.172122905.57636646/v1

10. Gorbach A.N., Tseitlin N.A. *Purchasing Behavior: Analysis of Spontaneous Sequence and Regression Models in Marketing Research*. Kiev, 2011. 220 p.

11. Bazilevskiy M.P. Construction of Quite Interpretable Linear Regression Models Using the Method of Successive Increase the Absolute Contributions of Variables to the General Determination. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Sistemnyi analiz i informatsionnye tekhnologii = Proceedings of Voronezh State University. Series: Systems Analysis and Information Technologies*, 2022, no. 2, pp. 5–16. (In Russian). EDN: CNDSSW. DOI: 10.17308/sait/1995-5499/2022/2/5-16.

12. Bazilevskiy M.P. Program for Constructing Quite Interpretable Elementary and Non-Elementary Quasi-Linear Regression Models. *Trudy Instituta sistemnogo programmirovaniya RAN = Proceedings of the Institute for System Programming of the RAS*, 2023, vol. 35, no. 4, pp. 129–144. (In Russian). EDN: KTOSCW. DOI: 10.15514/ISPRAS-2023-35(4)-7.

13. Bazilevskiy M.P. On the Solvability of the Optimization Problem for Constructing Quite Interpretable Linear Regressions. *Uchenye zapiski Kazanskogo universiteta. Seriya: Fiziko-matematicheskie nauki = Proceedings of Kazan University. Physical and Mathematical Series*, 2025, vol. 167, no. 4, pp. 627–640. (In Russian). EDN: PXUNAW. DOI: 10.26907/2541-7746.2025.4.627-640.

14. Bazilevskiy M.P. A Method for Constructing Fully Interpretable Linear Regressions with Statistically Significant Estimates According to the Student's Criterion and Insignificant Intercorrelation Coefficients. *Vychislitel'nye metody i programmirovanie = Numerical Methods and Programming*, 2025, vol. 26, no. 4, pp. 534–547. (In Russian). EDN: WKVQKH. DOI: 10.26089/NumMet.v26r435.

15. Optner S.L. *Systems Analysis for Business and Industrial Problem Solving*. New Jersey, Prentice Hall, 1965. 116 p. (Russ. ed.: Optner S.L. *Systems Analysis for Business and Industrial Problem Solving*. Moscow, Sovetskoe radio Publ., 1969. 216 p.).

16. Joung S. *Management: A Systems Analysis*. Foresman and Company, 1966. 436 p. (Russ. ed.: Joung S. *Management: A Systems Analysis*. Moscow, Sovetskoe radio Publ., 1972. 456 p.).

17. Fedorenko N.P. On Methods of Socio-Economic Forecasting. In *Methodology of Forecasting the Economic Development of the USSR*. Moscow, 1971, pp. 9–21. (In Russian).

18. Chernyak Yu.I. *Systems Analysis in Economic Management*. Moscow, Ekonomika Publ., 1975. 191 p.

19. Peregudov F.I. *Fundamentals of Systems Design of Automated Control Systems for Organizational Complexes.* Tomsk University Publ., 1984. 177 p.

20. Aivazyan S.A., Mkhitaryan V.S. *Applied Statistics and Essentials of Econometrics*. Moscow, YUNITI Publ., 1998. 1003 p.

Информация об авторе

Базилевский Михаил Павлович — кандидат технических наук, доцент, кафедра математики, Иркутский государственный университет путей сообщения, 664074, Российская Федерация, г. Иркутск, ул. Чернышевского, 15, e-mail: mik2178@yandex.ru, https://orcid.org/0000-0002-3253-5697, SPIN-код: 4347-5028.

Information about the Author

Mikhail P. Bazilevskiy — PhD in Technical Sciences, Associate Professor, Department of Mathematics, Irkutsk State Transport University, 15 Chernyshevsky Str., Irkutsk, 664074, Russian Federation, e-mail: mik2178@yandex.ru, https://orcid.org/0000-0002-3253-5697, SPIN-Code: 4347-5028.