ОБЪЯСНИМОСТЬ МОДЕЛЕЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ОЦЕНКЕ КРЕДИТОСПОСОБНОСТИ МАЛОГО БИЗНЕСА

Хаёт Хаитов
независимый исследователь
Сингапурский институт развития менеджмента в Ташкенте

Аннотация. *В тезисах рассматривается значение объяснимости моделей искусственного интеллекта при оценке кредитоспособности малого бизнеса. На основе анализа международных методологических материалов выделены основные требования к представлению и проверке результатов автоматизированной оценки. Предложен порядок взаимодействия модели, кредитного специалиста и предпринимателя. Особое внимание уделено достоверности исходных данных, понятности причин оценки и сохранению человеческого контроля. Сформулированы практические предложения по документированию результатов и рассмотрению спорных случаев.*

Ключевые слова: *Искусственный интеллект, малый бизнес, кредитоспособность, объяснимость, кредитный скоринг, модельный риск, человеческий контроль.*

ВВЕДЕНИЕ

Применение искусственного интеллекта в финансовой сфере расширяет возможности обработки информации и формирования прогнозов. Вместе с тем использование сложных моделей создаёт проблему понимания полученных результатов. Исследование Института финансовой стабильности показывает, что ограниченная объяснимость таких моделей затрудняет управление рисками и обеспечение подотчётности финансовых организаций.

В кредитовании малого бизнеса важно не только получить оценку риска, но и установить, какие сведения повлияли на неё. Кредитному специалисту необходимо проверить исходные данные, а предпринимателю — понять, какие обстоятельства были учтены при рассмотрении заявки. Формулировка «решение принято системой» не раскрывает содержания проведённого анализа.

Актуальность темы определяется необходимостью совместить использование автоматизированной оценки с возможностью проверки её оснований. Цель исследования — предложить простой порядок представления объяснений при оценке кредитоспособности малого бизнеса и определить роль специалиста в работе с результатами модели.

МЕТОД ИССЛЕДОВАНИЯ

В работе использованы анализ научно-методических материалов, сравнение и обобщение. Основу исследования составили документы Национального института стандартов и технологий США и Института финансовой стабильности. Предложенный порядок раскрывается на условном примере кредитной заявки. Реальные данные заёмщиков и результаты обучения моделей не использовались, поэтому выводы имеют концептуальный характер.

АНАЛИЗ ИССЛЕДОВАНИЯ

Объяснимость в рамках настоящего исследования рассматривается как возможность представить основания результата модели в форме, доступной соответствующему пользователю. При этом техническое описание алгоритма и понятное объяснение конкретной оценки решают разные задачи. NIST связывает объяснимость с представлением механизмов работы системы, а интерпретируемость — с пониманием значения её результата в контексте применения.

Для кредитного анализа предлагается выделить три взаимосвязанных элемента: проверку исходных сведений, представление основных факторов оценки и документирование окончательного решения.

❖ Первый элемент — проверка данных. Объяснение не будет полезным, если оно построено на ошибочных или неполных сведениях. Например, снижение поступлений может требовать проверки сопоставимости периодов, а отсутствие операций на одном счёте — уточнения того, использует ли предприятие другие расчётные каналы. В предлагаемом подходе необходимо фиксировать источник информации, дату её получения и выявленные ограничения.

❖ Второй элемент — представление факторов оценки. Специалисту нужен не только итоговый балл, но и информация о показателях, существенно повлиявших на результат. Предпринимателю те же обстоятельства следует сообщать понятным языком, без избыточных технических терминов.

❖ При этом перечень важных факторов нельзя автоматически считать достоверным объяснением. В исследовании Института финансовой стабильности отмечается, что методы объяснимости могут быть неточными, нестабильными и создавать вводящие в заблуждение объяснения. Следовательно, проверке подлежит не только прогноз модели, но и способ его интерпретации.

❖ Третий элемент — разграничение результата модели и кредитного решения. В предлагаемом порядке модель формирует оценку в пределах своей задачи, а окончательное решение учитывает также проверенные документы, действующие обязательства и условия запрашиваемого финансирования.

Поэтому в кредитном заключении целесообразно отдельно отражать машинную оценку и дополнительные основания решения.

Таблица 1

Предлагаемые требования к объяснению кредитной оценки

Требование	Практическое содержание
Понятность	Основные причины оценки изложены языком, доступным пользователю
Проверяемость	Для существенных показателей можно установить источник и период формирования
Соответствие модели	Объяснение отражает поведение используемого алгоритма
Разграничение оснований	Результат модели отделён от дополнительных причин кредитного решения
Возможность пересмотра	Предусмотрен порядок уточнения данных и рассмотрения спорного результата

Источник: разработано автором на основе обобщения рассмотренных материалов.

Рассмотрим условную ситуацию. Модель оценила заявку торгового предприятия как рискованную, указав на снижение поступлений за последние месяцы. При проверке специалист установил, что часть продаж была переведена на другой расчётный счёт. В таком случае необходимо уточнить полноту данных и при наличии оснований повторить оценку. Простое ручное изменение результата без фиксации причины не позволяет понять, была ли устранена ошибка или проигнорирован значимый сигнал.

Возможна и другая ситуация: снижение поступлений подтверждается всеми доступными источниками. Тогда объяснение помогает определить направление дальнейшего анализа, но само по себе не доказывает причину ухудшения положения предприятия.

Человеческий контроль в предлагаемом подходе означает возможность содержательной проверки, а не формального подтверждения вывода системы. Специалисту необходимы доступ к использованным сведениям и полномочия направить заявку на дополнительное рассмотрение. NIST подчёркивает значение человеческого суждения при управлении рисками AI-систем и допускает необходимость вмешательства человека, когда система не способна обнаружить или исправить ошибку.

Для практической организации работы предлагается сохранять краткую карточку оценки: дату расчёта, версию модели, основные факторы, ограничения данных и результат проверки специалистом. Такой документ позволит

восстановить последовательность рассмотрения заявки и проанализировать спорные случаи.

ЗАКЛЮЧЕНИЕ

Объяснимость моделей искусственного интеллекта следует рассматривать как условие проверяемого использования автоматизированной оценки, а не как дополнительный комментарий к итоговому баллу. Её назначение заключается в том, чтобы показать основания результата и обозначить пределы его применения.

В работе предложены три направления организации объяснений: проверка исходных данных, представление существенных факторов и разграничение модельной оценки с окончательным кредитным решением.

Для банковской практики целесообразно предусмотреть краткое документирование результатов, порядок исправления ошибочных сведений и дополнительное рассмотрение спорных заявок. Эффективность этих мер необходимо проверять на реальных данных — по способности выявлять ошибки и повышать понятность кредитного анализа.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. National Institute of Standards and Technology (NIST). *Artificial Intelligence Risk Management Framework (AI RMF 1.0).* NIST AI 100-1. 2023. DOI: 10.6028/NIST.AI.100-1.
2. Perez-Cruz, F., Prenio, J., Restoy, F., Yong, J. *Managing explanations: How regulators can address AI explainability.* FSI Occasional Papers. No. 24. Bank for International Settlements, 2025.
3. Ribeiro, M. T., Singh, S., Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. P. 1135–1144. DOI: 10.1145/2939672.2939778.
4. Lundberg, S. M., Lee, S.-I. A Unified Approach to Interpreting Model Predictions // Advances in Neural Information Processing Systems. 2017. Vol. 30. P. 4765–4774.