

Секция «Обеспечение экономической безопасности государства в условиях современных глобальных вызовов: экономико-правовые аспекты (совместно с Экономическим факультетом Белорусского государственного университета)»

Использование машинного обучения в скоринговых системах: новое поколение оценки рисков

Научный руководитель – Маслюкова Татьяна Ивановна

Григорьев Константин Дмитриевич

Студент (бакалавр)

Белорусский государственный университет, Экономический факультет, Минск, Беларусь

E-mail: grigoryev.konstantin15@gmail.com

Введение. Современные скоринговые системы играют важную роль в сфере финансовых технологий, давно перешагнув пределы классических статистических моделей. Сегодня они представляют собой мощные инструменты анализа рисков, позволяя не только оценивать платежеспособность заемщика, но и выявлять потенциальные угрозы для бизнеса на самых ранних этапах. В условиях нестабильности финансовых рынков и растущей неопределенности применение технологий машинного обучения (ML) в анализе рисков становится ключевым элементом для банков и других финансовых структур. Использование ML расширяет возможности традиционных методов, позволяя учитывать больше факторов, включая клиентское поведение и макроэкономические тенденции. В данной статье мы рассмотрим, как машинное обучение трансформирует подход к оценке рисков и почему его применение становится критически важным для принятия обоснованных финансовых решений.

Основная часть. Скоринговая система — это аналитический инструмент, предназначенный для поддержки принятия решений на основе данных. Она применяется в разных сферах: от кредитного одобрения до оценки благонадежности клиентов и прогнозирования рисков. Классические модели, использующие ограниченные объемы данных, не всегда учитывают важные факторы, например, отсутствие или малый объем кредитной истории, что делает определенные группы заемщиков "невидимыми" для подобных систем. В США, например, более 45 миллионов граждан не имеют полноценной кредитной истории, а в ряде стран эта цифра еще выше. В Индии к таким "невидимкам" относится около 63% населения, в ЮАР — 51%, а в Гонконге — 16%. Поскольку традиционные системы оценивают заемщиков исключительно на основе их кредитного прошлого, значительная часть потенциальных клиентов остается вне поля зрения, что снижает эффективность анализа [3].

Использование ML-моделей позволяет работать с большим объемом данных, включая демографические показатели, финансовые операции, социальную активность и макроэкономическую обстановку, что минимизирует операционные риски и повышает точность прогнозирования. Например, если уровень инфляции или ключевая ставка резко изменяются, такие факторы также должны учитываться при анализе кредитных рисков. Таким образом, применение машинного обучения в скоринговых системах не только автоматизирует процесс оценки, но и повышает его точность, снижая вероятность ошибок.

Работа скоринговой модели начинается со сбора данных о клиенте: его демографических характеристик, финансового положения, истории займов и других параметров. Качество итогового анализа во многом зависит от полноты и точности собранной информации. После этого данные проходят этап предобработки, включающий заполнение пропущенных значений, устранение выбросов и удаление нерелевантных данных. Затем модель обучается с использованием различных алгоритмов, таких как логистическая регрессия, деревья

решений или градиентный бустинг. Эти методы позволяют строить классификационные модели, определяющие вероятность дефолта заемщика.

При создании скоринговых моделей активно используются мощные библиотеки машинного обучения, включая Scikit-learn, XGBoost, LightGBM и CatBoost [1]. Каждая из них предоставляет инструменты для разработки высокоточных предсказательных моделей. Scikit-learn поддерживает широкий спектр алгоритмов, включая логистическую регрессию, деревья решений и случайный лес, которые являются популярными инструментами классификации. XGBoost и LightGBM, основанные на градиентном бустинге, позволяют строить более сложные и точные прогнозные модели, что особенно важно при оценке кредитных рисков.

Простейшим примером алгоритма является дерево решений [2]. В рамках его работы на первом этапе может происходить разделение клиентов по уровню дохода. Если доход превышает определенное значение, заемщика относят к категории "низкого риска". Далее анализируется платежная дисциплина: если у клиента более трех просрочек, его перемещают в группу "высокого риска", даже если он имеет высокий доход.

На рисунке 1 представлено более детальное дерево решений, применяемое в кредитном скоринге. Данный алгоритм позволяет принимать решения о выдаче кредита на основе нескольких ключевых факторов:

- 1) **Возраст заемщика** – если он **старше 40 лет**, проверяются дополнительные параметры, если **моложе** – проводится оценка кредитной истории.
- 2) **Цель кредитования** – различаются две категории: "Открытие бизнеса" и "Повседневные расходы".
 - Если заемщик берет кредит для бизнеса, анализируется наличие **бизнес-плана**. Если он есть, кредит одобряется, если нет – следует отказ.
 - Если средства запрашиваются на повседневные нужды, кредит предоставляется без дополнительных условий.
 - **Кредитная история** – при наличии **отличной кредитной истории** кредит одобряется. Если история оставляет желать лучшего, проверяется наличие **имущества для залога**.
 - При наличии залога кредит одобряется, если его нет – заявка отклоняется.

Заключение. Внедрение машинного обучения в процесс скоринга значительно улучшает способность финансовых организаций более точно анализировать риски и принимать взвешенные решения. ML-модели позволяют учитывать больше факторов, адаптироваться к изменениям условий и формировать прогнозы на основе больших массивов данных. Однако, несмотря на технологический прогресс, остаются нерешенные вопросы, например, невозможность полностью учесть контекст ситуации или уязвимость перед мошенническими схемами. Тем не менее, машинное обучение продолжает оказывать значительное влияние на процессы оценки рисков, делая их более точными и гибкими. В дальнейшем стоит ожидать еще более глубокого внедрения ML в управление рисками и появления новых, более совершенных инструментов для финансового анализа.

Источники и литература

- 1) Воронов, А. В. Машинное обучение: учебное пособие / А. В. Воронов. – М.: Юрайт, 2020. – 256 с.

- 2) Что такое дерево решений и где его используют? [Электронный ресурс] // Хабр. Сообщество IT специалистов. URL: <https://habr.com/ru/companies/productstar/article/523044/> (дата обращения: 10.01.2025).
- 3) TransUnion. More than 45 million Americans are either credit unserved or underserved — Approximately 20% migrate to being credit active every two years. 2025. [Электронный ресурс] // Newsroom.transunion. URL: https://newsroom.transunion.com/more-than-45-million-americans-are-either-credit-unserved-or-underserved—approximately-20-migrate-to-being-credit-active-every-two-years#_edn1 (дата обращения: 16.01.2025).

Иллюстрации

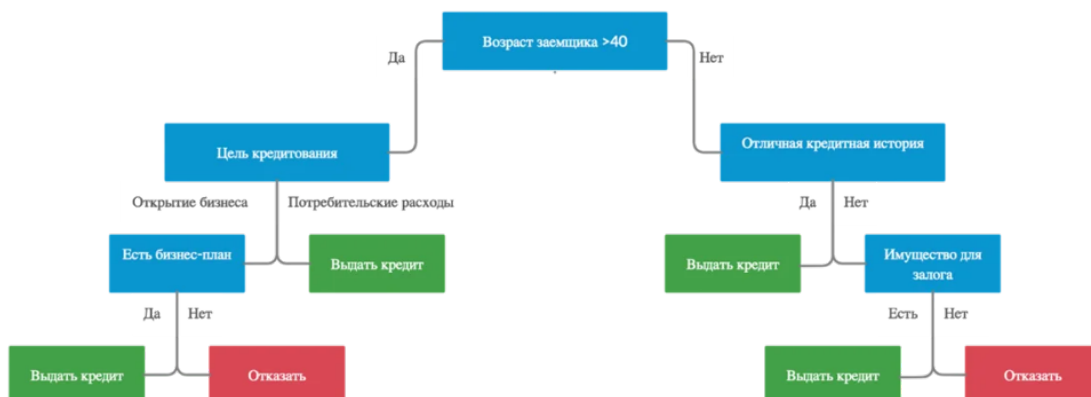


Рис. : Пример дерева решений для оценки кредитного риска