

ПРОГНОЗИРОВАНИЕ КРЕДИТОСПОСОБНОСТИ С ПОМОЩЬЮ КРЕДИТНОГО СКОРИНГА С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

PREDICTING CREDITWORTHINESS USING CREDIT SCORING USING MACHINE LEARNING METHODS

A. Karmazin

Summary. The study of creditworthiness using machine learning methods is not only of academic interest, but also has direct practical significance for financial institutions, insurance companies and other market participants. The use of advanced machine learning techniques provides an opportunity not only for more accurate forecasting, but also for deeper data analysis, which, in turn, helps identify early signs of financial risks. Current trends in scoring include the use of digital technologies, including artificial intelligence, machine learning, and automation of scoring processes. At the same time, an integrated approach combining subjective and statistical approaches to scoring, as well as taking into account specific factors of the factoring business, has become widespread. Thus, the use of machine learning methods in credit scoring can significantly improve the accuracy of predicting borrowers' creditworthiness. However, it is necessary to regularly update models and maintain a balance between the complexity of the model and its interpretability. Choosing the right algorithm and carefully preparing the data are key success factors in building an effective scoring system for predicting borrowers' creditworthiness.

Keywords: scoring, machine learning, lending, forecasting, financial organizations.

Кармазин Антон Ричардович

Аспирант, Российский университет дружбы народов
имени Патриса Лумумбы
karmaz.2010@yandex.ru

Аннотация. Исследование кредитоспособности с использованием методов машинного обучения представляет не только академический интерес, но и имеет непосредственное прикладное значение для финансовых учреждений, страховых компаний и других участников рынка. Использование передовых методов машинного обучения предоставляет возможность не только для более точного прогнозирования, но и для более глубокого анализа данных, что, в свою очередь, помогает выявлять ранние признаки финансовых рисков. Современными тенденциями в скоринге является использование цифровых технологий, в том числе искусственного интеллекта, применение машинного обучения, а также автоматизация скоринговых процессов. При этом, широкое распространение получил комплексный подход сочетающий субъективные и статистические подходы к скорингу, а также учёт специфических аспектов факторингового бизнеса. Таким образом, использование методов машинного обучения в кредитном скоринге позволяет значительно повысить точность прогнозирования кредитоспособности заёмщиков. Однако необходимо регулярное обновление моделей и соблюдение баланса между сложностью модели и её интерпретируемостью. Правильный выбор алгоритма и тщательная подготовка данных являются ключевыми факторами успеха в построении эффективной скоринговой системы прогнозирования кредитоспособности заемщиков.

Ключевые слова: скоринг, машинное обучение, кредитование, прогнозирование, финансовые организации.

Введение

В современном финансовом мире, где конкуренция на рынке кредитования постоянно растет, актуальность разработки эффективных методов прогнозирования кредитоспособности высока [1]. Точность и надежность таких методов являются ключевыми факторами для финансовых учреждений, стремящихся минимизировать риски и обеспечить устойчивость своих кредитных портфелей [2].

Исследование кредитоспособности с использованием методов машинного обучения представляет не только академический интерес, но и имеет непосредственное прикладное значение для финансовых учреждений, страховых компаний и других участников рынка. Результаты подобных исследований могут послужить основой

для оптимизации процессов принятия решений в сфере кредитования и обеспечить более эффективное управление рисками [3].

Актуальность темы прогнозирования кредитоспособности по кредитному скорингу с использованием методов машинного обучения трудно переоценить в свете современных вызовов в условиях растущего объема данных и разнообразия факторов, влияющих на финансовое положение заёмщиков, так как стандартные методы оценки кредитоспособности оказываются недостаточно эффективными. Использование передовых методов машинного обучения предоставляет возможность не только для более точного прогнозирования, но и для более глубокого анализа данных, что, в свою очередь, помогает выявлять ранние признаки финансовых рисков [4].

Целью работы является анализ инструментов прогнозирования кредитоспособности с помощью кредитного скоринга с использованием методов машинного обучения.

Методы

Данное исследование сочетает в себе взаимодополняющие общенаучные теоретические методы, включая анализ, синтез и абстрагирование. Интеграция анализа и синтеза обеспечивает всестороннее изучение аспектов прогнозирования кредитоспособности. Посредством синтеза отдельные компоненты и свойства, выявленные в ходе анализа, объединяются.

Метод абстракции используется для получения обобщений и выявления тенденций прогнозирования кредитоспособности с помощью кредитного скоринга с применением методов машинного обучения.

Результаты

Кредитный скоринг представляет собой систему оценки заемщика, которая позволяет кредитным организациям сформировать представление о потенциальном клиенте, а также спрогнозировать его поведение при выплате кредита [5].

Одним из основных видов скоринга является анкетный скоринг, позволяющий оценить кредитоспособность при первичной заявке, учитывающий возраст, доход, занятость и семейное положение заемщика, что определяет широкую распространённость данного вида скоринга. Скоринг поведения опирается на оценку заемщика на основании анализа истории платежей, характера трат, использования кредитных продуктов и применяется для действующих клиентов. Скоринг, направленный на противодействие мошенничеству, позволяет оценить риск мошеннических действий на основе сверки с базами данных и используется совместно с другими видами скоринга [6]. Скоринговые модели могут отличаться, но все они направлены на комплексную оценку потенциально заемщика и минимизацию рисков невозврата кредита.

Распространённой моделью кредитного скоринга при факторинге является анализ истории платежей (своевременность внесения, наличие просрочек, качество кредитной истории), анализ использования кредита (процент использования доступного кредитного лимита), длительность кредитной истории (срок существования кредитных счетов, стабильность кредитной истории), типы кредитных счетов (разнообразие кредитных продуктов, наличие факторинговых операций), а также анализ последних запросов (частота обращения за кредитом, количество запросов за последнее время) [7].

К преимуществам использования данных моделей относится объективность оценки, заключающаяся в ис-

ключении человеческого фактора, а также стандартизация процесса оценки. Оперативность принятия решений о финансировании обеспечивается за счёт автоматизации процесса оценки, что позволяет также снизить риски на основе прогнозирования вероятности невозврата средств и выявления мошеннических схем.

Современной тенденцией в скоринге является использование цифровых технологий, в том числе искусственного интеллекта, применение машинного обучения, а также автоматизация скоринговых процессов. При этом широкое распространение получил комплексный подход, сочетающий субъективные и статистические подходы к скорингу, а также учёт специфических аспектов факторингового бизнеса.

Инновационные методы скоринга позволяют финансовым организациям более точно оценивать риски, принимать обоснованные решения и расширять возможности кредитования для различных категорий заемщиков, что особенно важно в условиях растущей конкуренции на финансовом рынке. Так, финтех-компания LenddoEFL анализирует цифровой след и применяет машинное обучение для оценки надёжности заемщика. Компания ZestFinance использует алгоритмы машинного обучения для определения кредитоспособности заемщика [8].

Одним из основных методов скоринга является традиционный скоринг, основанный на финансовых показателях с использованием стандартных коэффициентов, при этом он требует наличия полной финансовой отчётности и зависит от качества изначально предоставленных данных. Статистический скоринг использует математические модели, анализирует большие массивы данных и учитывает исторические показатели, а также применяет методы регрессионного анализа. Кредитный скоринг позволяет оценивать платёжеспособность клиента, анализирует кредитную историю, учитывает финансовое положение, а также позволяет оценить поведение заемщика. Поведенческий скоринг позволяет анализировать поведение клиента, оценивает паттерны платежей, учитывает нефинансовые показатели, а также использует данные о взаимодействиях клиента. Альтернативный скоринг использует нетипичные данные для анализа, в том числе цифровой след, социальные связи клиента и учитывает поведенческие факторы. Результаты проведённого сравнительного анализа эффективности различных методов скоринга приведены в таблице 1.

Таким образом, по результатам анализа таблицы 1, традиционный скоринг может быть отнесён к оптимальному варианту использования для крупных компаний, которые характеризуются наличием полной отчётности. Статистический скоринг наиболее подходит для массового факторинга. Кредитный скоринг наиболее применим для оценки истории платежей. Поведенческий скоринг позволяет полнее оценить динамику взаимо-

Таблица 1.

Результаты проведённого сравнительного анализа эффективности различных методов скоринга

| Критерий | Традиционный | Статистический | Кредитный | Поведенческий | Альтернативный |
|------------------------------|--------------|----------------|-----------|---------------|----------------|
| Скорость обработки | Низкая | Высокая | Средняя | Высокая | Высокая |
| Точность оценки | Средняя | Высокая | Высокая | Высокая | Высокая |
| Зависимость от данных | Высокая | Средняя | Средняя | Низкая | Низкая |
| Стоимость внедрения | Низкая | Высокая | Средняя | Высокая | Высокая |
| Адаптивность | Низкая | Средняя | Средняя | Высокая | Высокая |
| Устойчивость к мошенничеству | Низкая | Средняя | Высокая | Высокая | Высокая |

действия. Альтернативный скоринг является наиболее предпочтительным для анализа новых клиентов без истории кредитных отношений.

Также, при осуществлении скоринга целесообразно использовать комплексный подход, заключающийся в анализе базовых финансовых показателей, статистическом анализе, проведении поведенческого анализа, использовании альтернативных источников данных, а также применении машинного обучения для прогнозирования поведения клиентов. Этот подход позволит повысить точность оценки кредитоспособности, снизить риски невозврата средств, ускорить процесс принятия решений, расширить возможности оценки новых клиентов и адаптировать систему под меняющиеся условия рынка. При этом, внедрение инновационных методов скоринга требует значительных инвестиций, но обеспечивает существенное повышение эффективности оценки кредитоспособности и снижение рисков невозврата средств.

Основными этапами построения скоринговой модели на основе машинного обучения являются: сбор и подготовка данных (сбор исторических данных о заемщике, очистка данных от пропусков и выбросов, преобразование категориальных переменных, нормализация числовых показателей); выбор признаков (анализ корреляции между признаками, отбор более информативных переменных, создание новых признаков); построение модели (выбор алгоритма машинного обучения, разделение данных на обучающую и тестовую выборки, обучение моделей, валидация результатов).

В итоге, скоринговые модели выводят числовой балл, который количественно определяет кредитоспособность заемщика. Этот балл является непрерывной переменной, получаемой в результате агрегирования и преобразования входных характеристик с помощью алгоритмов машинного обучения [9].

Регрессия с опорными векторами (SVR) относится к категории непараметрических методов из-за её зависимости от функций ядра [10]. Регрессия с опорными

векторами — это форма машинного обучения, которая фокусируется на прогнозировании непрерывной переменной результата на основе набора переменных-предикторов. Метод SVR основан на принципах метода опорных векторов, который чаще всего используется для задач классификации. Основная цель SVR состоит в том, чтобы найти функцию, которая отклоняется от фактических наблюдаемых результатов на величину, не превышающую заданного порогового значения. Использование ядер в SVR позволяет выполнять нелинейную регрессию, неявно отображая входные данные в виде пространственных объектов более высокой размерности, где выполняется линейная регрессия. Эта возможность делает SVR универсальным и применимым к различным задачам регрессии, где взаимосвязь между входными и выходными переменными не является линейной.

Метод CatBoost относится к семейству деревьев принятия решений (GBDT), он позволяет напрямую обрабатывать категориальные переменные и использует градиентное ускорение для деревьев принятия решений. Одной из ключевых особенностей CatBoost является инновационный подход к кодированию категорий. Таким образом, CatBoost эффективно обрабатывает категориальные данные в задачах классификации и регрессии. Алгоритм также использует схему для вычисления конечных значений при выборе древовидной структуры, что позволяет избежать переобучения [11].

LightGBM — это платформа для повышения градиента, характеризующаяся скоростью и эффективностью. Это связано с такими методами как односторонняя выборка на основе градиента и эксклюзивное объединение функций. LightGBM обеспечивает повышенную эффективность за счет снижения требований к размеру данных для точных прогнозов и уменьшения количества функций модели без снижения точности [12].

Elastic net — это метод регуляризации, который эффективно решает проблемы мультиколлинеарности и выбора переменных в наборах данных [13]. Этот метод

позволяет сократить количество переменных в модели за счет выполнения выбора переменных, а также эффективно работает с моделями, где независимые переменные сильно коррелируют.

К метрикам оценки качества модели относятся: общая точность модели (Accuracy); точность положительных прогнозов (Precision); полнота обнаружения положительных случаев (Recall); гармоническое среднее между precision и recall (F1-score); площадь под ROC-кривой (AUC-ROC).

При прогнозировании кредитоспособности с использованием инструментов машинного обучения используются следующие типы признаков: демографические (возраст, пол, семейное положение); финансовые (доход, кредитная история, долговая нагрузка); поведенческие (активность в банковских приложениях, история транзакций). При этом могут возникать проблемы дисбаланса классов, которые могут быть устранены за счёт синтетического увеличения данных (SMOTE), использования взвешенных алгоритмов и срезания доминирующего класса для выравнивания дисбаланса классов.

Для оценки редких категорий может быть использован Target Encoding, или Embedding-слои в нейросетях для категориальных данных. Анализ временных признаков основывается на учёте сезонности (например, кредиты перед праздниками) и скользящих окон для анализа истории платежей. К современным алгоритмам и фреймворкам относятся LightGBM и CatBoost, которые эффективны для анализа реальных данных и имеют встроенную обработку пропусков в данных [8].

Интерпретация полученных моделей по итогам использования методов машинного обучения заключается в визуализации вклада каждого анализируемого признака в прогноз с использованием инструмента SHAP (SHapley Additive exPlanations).

Этапы внедрения инструментов машинного обучения в скоринговый прогноз состоят из проверки данных на примере признаков, связанных с целевой переменной, тестировании модели на разных сегментах заёмщиков, интеграции с CRM-системой банка, а также регулярной корректировки модели.

Заключение

Инновационные методы скоринга позволяют финансовым организациям более точно оценивать риски, принимать обоснованные решения и расширять возможности кредитования для различных категорий заёмщиков, что особенно важно в условиях растущей конкуренции на финансовом рынке. Таким образом, использование методов машинного обучения в кредитном скоринге позволяет значительно повысить точность прогнозирования кредитоспособности заёмщиков. Однако необходимо регулярное обновление моделей и соблюдение баланса между сложностью модели и её интерпретируемостью. Правильный выбор алгоритма и тщательная подготовка данных являются ключевыми факторами успеха в построении эффективной скоринговой системы прогнозирования кредитоспособности заёмщиков.

ЛИТЕРАТУРА

1. Заболоцкая В.В. Методы диагностики и прогнозирования кредитоспособности субъектов МСП с применением искусственного интеллекта // Известия Саратовского университета. Серия: Экономика. Управление. Право. — 2024. — № 3. — С. 294–311.
2. Abdou H.A., Pointon J. Credit scoring, statistical techniques, and evaluation criteria: A review of the literature // Intelligent Systems in Accounting, Finance & Management. — 2011. — № 12. — pp. 59–88.
3. Луценко Е.В., Коваленко А.В. Открытая персональная интеллектуальная технология разработки и применения адаптивных методик оценки инвестиционной привлекательности и кредитоспособности предприятий // Вестник Пермского университета. Сер. «Экономика». — 2019. — № 1. — С. 20–50.
4. Xiao-Lin L., Zhong Yu. An Overview of Personal Credit Scoring: Techniques and Future Work // International Journal of Intelligence Science. — 2012. — № 2. — pp. 181–189.
5. Носова Т.П., Паршин А.Б., Терпицкая К.И. Классификация банковских рисков и мероприятия по их снижению с целью оптимизации банковской деятельности // Вестник Академии знаний. — 2022. — № 6. — С. 349–354.
6. Поляков К.Л., Жукова Л.В. Опыт моделирования вероятности кредитного дефолта клиентов микрофинансовых организаций // Экономический журнал Высшей школы экономики. — 2019. — № 4. — С. 497–523.
7. Шевелев А.А., Бузанов Г.Н. Модель вероятности дефолта с использованием транзакционных данных российских компаний // Серия докладов об экономических исследованиях Банка России. — 2022. — № 9. — С. 9–19.
8. Alamsyah A., Hafidh A. A., Mulya A. D. Innovative Credit Risk Assessment: Leveraging Social Media Data for Inclusive Credit Scoring in Fintech Sector // Journal of Risk and Financial Management. — 2025. — № 2. — pp. 74–91.
9. Инхиреева Т.И. Методология подготовки исходных данных для построения кредитного скоринга // Информационные технологии в науке, управлении, социальной сфере и медицине. — 2019. — С. 246–251.
10. Kectman V., Kopriva I. Kernel Based Algorithms for Mining Huge Data Sets: Supervised, Semi-Supervised, and Unsupervised Learning. Berlin: Springer, 2006. — 276 p.
11. Hancock J.T., Khoshgoftaar T.M. CatBoost for big data: An interdisciplinary review // Journal of Big Data. — 2020. — № 1. — pp. 94–107.
12. Sheridan R.P., Liaw A., Tudor M. Light Gradient Boosting Machine as a Regression Method for Quantitative Structure Activity Relationships // Plos One. — 2021. — № 7. — pp. 163–174.
13. Li J., Wang X., Yang X. Analyzing freeway safety influencing factors using the catboost model and interpretable machine-learning framework, SHAP // Transportation Research Record: Journal of the Transportation Research Board. — 2023. — № 7. — pp. 563–574.

© Кармазин Антон Ричардович (karmaz.2010@yandex.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»