

УДК 336.774.3

ПРОЦЕСС РАЗРАБОТКИ МОДЕЛЕЙ КРЕДИТНОГО СКОРИНГА В КРЕДИТНОЙ ОРГАНИЗАЦИИ НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Казакова Анна Константиновна,

студент, кафедра прикладной математики
института информационных технологий
МИРЭА – Российского технологического университета,
г. Москва
E-mail: messageann@yandex.ru

Царёв Роман Юрьевич,

канд. техн. наук, доцент кафедры прикладной математики
института информационных технологий
МИРЭА – Российского технологического университета,
г. Москва
E-mail: caryov@mirea.ru

Аннотация

В статье рассматривается процесс разработки скоринговых моделей в кредитных организациях с применением методов машинного обучения как современный неотъемлемый элемент системы управления кредитными рисками. Обосновывается актуальность использования методов машинного обучения для повышения точности прогнозирования и устойчивости моделей. Обосновывается необходимость системного подхода к организации процесса такой разработки, подчёркивается значение такого подхода в условиях высокой сложности и ресурсоёмкости проектов скоринга в крупных финансовых организациях. Посредством моделирования бизнес-процесса в нотации IDEF0 в статье представлен жизненный цикл разработки скоринговой модели, включая описание и декомпозицию отдельных ключевых этапов: от постановки задачи до внедрения в производственные процессы принятия решений.

Ключевые слова: управление рисками, кредитный скоринг, скоринговая модель, машинное обучение, моделирование бизнес-процессов, разработка моделей машинного обучения.

THE PROCESS OF DEVELOPING CREDIT SCORING MODELS IN CREDIT INSTITUTIONS USING MACHINE LEARNING METHODS

Kazakova Anna Konstantinovna,

Student, Applied Mathematics Department,
Institute of Information Technologies,
MIREA – Russian Technological University,

Russia, Moscow

E-mail: messageann@yandex.ru

Tsarev Roman Yurievich,

Ph.D. of Engineering Sciences,

Associate Professor of the Applied Mathematics Department,

Institute of Information Technologies,

MIREA – Russian Technological University,

Russia, Moscow

E-mail: caryov@mirea.ru

ABSTRACT

The article explores the process of developing credit scoring models in credit institutions using machine learning methods as a modern and essential component of credit risk management systems. It substantiates the relevance of applying machine learning techniques to enhance accuracy and model stability. The paper highlights the necessity of a systematic approach to organize the development process, emphasizing its importance in the context of high complexity and resource intensity of scoring projects in large financial institutions. Through business process modeling in IDEF0 notation, the article presents the complete life cycle of credit scoring model development, including a structured description and decomposition of its key stages – from problem formulation to implementation in production decision-making processes.

Keywords: risk management, credit scoring, scoring model, machine learning, business process modeling, machine learning model development.

Введение. В современных условиях развития кредитного рынка и усиления конкуренции банки и финансовые организации сталкиваются с необходимостью эффективного управления кредитными рисками. Вне зависимости от типа предоставляемого кредита – ипотека, автокредит, кредитная карта, – управление кредитными рисками является неотъемлемой частью деятельности любой кредитной организации, так как напрямую влияет на её финансовую устойчивость. Одной из ключевых задач в этом направлении является оценка платёжеспособности клиента или оценка вероятности его дефолта – неспособности заёмщика выполнить свои финансовые обязательства перед банком. Такой прогноз, основанный на исторических и актуальных данных, реализуется с помощью понятия кредитного скоринга, от точности которого зависит устойчивость кредитного и прибыльность деятельности кредитной организации в целом [1]. Кредитный скоринг является неотъемлемой частью системы управления рисками в банке. На основе его результатов могут приниматься решения о выдаче кредита, установлении кредитного лимита, корректировке условий обслуживания или применении превентивных мер по снижению убытков, ввиду чего надёжность скоринговых решений напрямую влияет на качество кредитного портфеля.

Для решения задачи традиционно применяются скоринговые модели, базирующиеся на экспертных оценках и современных методах статистического анализа и машинного обучения. С развитием технологий, ростом доступности данных и усложнением требований к качеству моделей становится необходимым формализованный подход к процессу их разработки, что особенно актуально при использовании методов

машинного обучения, где жизненный цикл модели включает множество итеративных этапов. Для повышения управляемости, эффективности и воспроизводимости процессов активно используется моделирование бизнес-процессов.

Цель исследования. Целью настоящей работы является описание и обоснование формализованного бизнес-процесса разработки моделей кредитного скоринга на основе методов машинного обучения для банковской сферы. Объектом исследования является процесс управления кредитными рисками в банке, поскольку именно в его рамках создаются скоринговые модели, предмет исследования – это бизнес-процесс разработки моделей кредитного скоринга с использованием методов машинного обучения.

Для достижения цели видится необходимым:

обосновать преимущества использования бизнес-процессного подхода при разработке моделей на основе машинного обучения;

проанализировать структуру жизненного цикла скоринговой модели и выделить ключевые этапы её разработки и внедрения;

смоделировать и представить формализованную схему рассматриваемого бизнес-процесса с подробным описанием;

рассмотреть особенности информационных потоков процесса исходя из стандартов в области машинного обучения.

Материалы и методы исследования. Скоринговая модель представляет собой инструмент, который на основе ряда факторов рассчитывает вероятность выполнения заемщиком своих обязательств, актуальных в исследуемый период жизни кредита. Исторически методы оценки кредитоспособности основывались на экспертных оценках и качественном анализе, но с развитием статистики и накоплением объёмов данных во второй половине XX века начали применяться и количественные методы, такие как линейная регрессия и дискриминантный анализ. В последние десятилетия, с развитием вычислительных мощностей и появлением перспективных методов машинного обучения, кредитный скоринг эволюционировал, включив в себя более сложные алгоритмы и модели. Такие модели обладают способностью выявлять сложные нелинейные зависимости в данных, учитывая множество факторов одновременно, чем обеспечивают более высокую точность прогнозов по сравнению с традиционными статистическими методами [2].

В наши дни реализовано уже множество моделей скоринга, использующих именно методы машинного обучения. Практика показывает, что их применение в кредитном скоринге является не просто исследовательским трендом, а уже признанным стандартом в индустрии. Особенно широко подобные решения используются в крупных банках, обладающих обширными историческими данными и ресурсами для внедрения высокотехнологичных инструментов принятия решений, в их числе:

Сбербанк: активно реализует скоринговые системы, основанные на ансамблевых моделях и нейронных сетях [3];

Т-Банк: использует нейросетевые скоринговые модели с антифрод-логикой, способные выявлять заявки на кредиты от заемщиков, находящихся под влиянием мошенников [4];

Банк Точка: описывает процесс разработки самой большой модели кредитного скоринга для малого и среднего бизнеса с помощью градиентного бустинга и построения ансамбля моделей [5];

Альфа-Банк: имеет опыт не просто построения систем на основе нейронных сетей и градиентного бустинга [6], а целые собственные разработки программных решений по автоматизации процесса создания и управления своими моделями [7].

Учитывая, что построение скоринговых моделей в кредитных организациях требует значительных массивов данных, участия разнопрофильных специалистов и соблюдения

нормативных требований, становится очевидной необходимостью четкого организационного подхода к управлению этим процессом. Особенно сложным становится процесс при использовании методов машинного обучения, которые предполагают более сложные циклы разработки по сравнению с традиционными статистическими моделями. Для достижения требуемой точности, устойчивости и интерпретируемости моделей необходимо пройти множество этапов и без формализованного подхода возрастает риск дублирования усилий, задержек, несогласованности, а также потери фокуса на цели и бизнес-ценности модели.

Моделирование бизнес-процесса позволяет систематизировать жизненный цикл разработки скоринговой модели: от постановки задачи до внедрения в процесс принятия решений. Такой подход не только упрощает управление сроками и ресурсами, но и снижает операционные и регуляторные риски [8]. Более того, следование четко определённой модели позволяет не потерять стратегический смысл проекта на протяжении всей разработки: каждая активность и каждое решение соотносятся с изначально поставленными целями модели и её назначением в системе управления рисками. Это особенно актуально для крупных организаций, где масштаб данных и сложность организационных структур требуют стандартизации и воспроизводимости процесса, а каждый вид скоринга предъявляет свои требования к архитектуре моделей, отбору признаков, периодичности пересчёта и способам верификации результатов.

Моделирование видится целесообразным проводить с использованием нотации IDEF0 – одного из стандартов функционального моделирования, который позволяет формализовать сложные бизнес-процессы в виде иерархической структуры функций и связей между ними. Эта нотация позволяет четко обозначить входы, выходы, механизмы и управляющие воздействия каждой стадии разработки, обеспечивая прозрачность, структурированность и согласованность всех этапов. Другим преимуществом является её универсальность и понятность для участников процесса с разным уровнем подготовки – от аналитиков до бизнес-руководителей [8].

Результаты и их обсуждение. Процесс разработки моделей кредитного скоринга представляется как комплексный и достаточно унифицированный бизнес-процесс, включающий в себя ряд этапов:

Постановка задачи – формулирование целей разработки модели, описание требований, выбор целевой переменной и регламентирование сроков и подходов к разработке.

Формирование выборки – определение критериев включения данных и сбор информации, релевантной поставленной задаче.

Подготовка данных – очистка, предварительный анализ, генерация признаков, формирование выборок разработки.

Отбор признаков и предварительное моделирование – выбор информативных и стабильных переменных, предварительное обучение модели, корректировка выделенных признаков при необходимости.

Обучение и оценка модели – построение итоговой модели на основе финального списка переменных, подбор оптимальных параметров, оценка качества по утверждённым метрикам, анализ интерпретируемости и устойчивости во времени.

Валидация – проверка соответствия модели требованиям регуляторов, внутренним стандартам и статистическим критериям.

Внедрение – интеграция модели в бизнес-процессы и использование её результатов в системе принятия решений. После внедрения обеспечивается мониторинг качества модели в продуктивной среде и, при необходимости, её переобучение.

Контекстная диаграмма моделируемого бизнес-процесса «Разработка модели кредитного скоринга в кредитной организации» представлена на Рисунке 1 и отражает все входные и выходные ресурсы, участвующие в процессе, а также его управляющие условия и механизмы. На Рисунке 2 приведена декомпозиция контекстной диаграммы, отражающая семь подпроцессов модели.

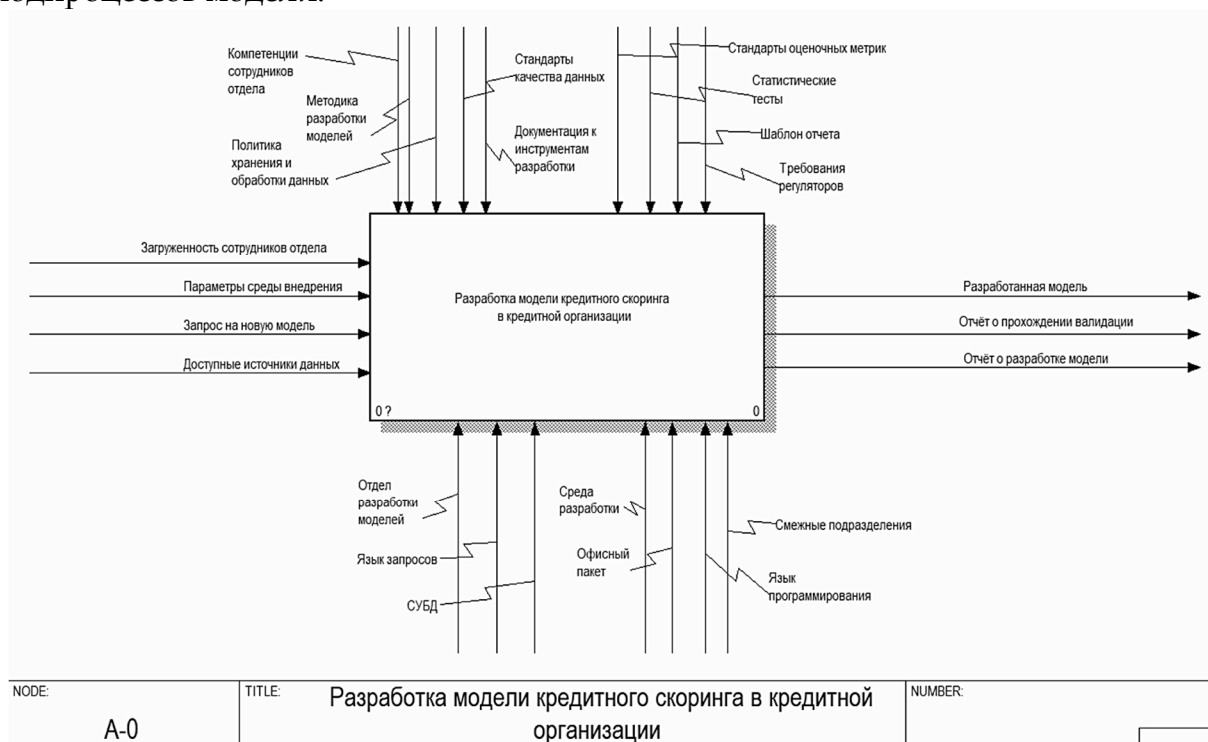


Рисунок 1 – Контекстная диаграмма процесса в нотации IDEF0

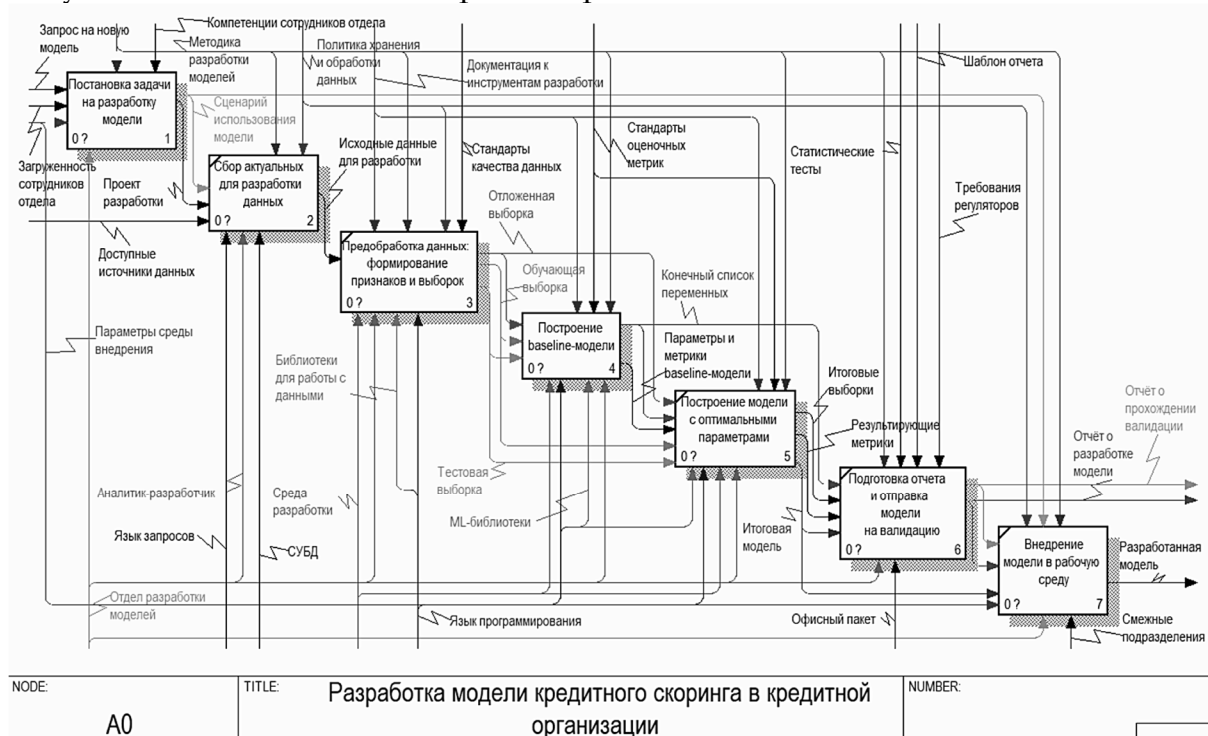


Рисунок 2 – Декомпозиция контекстной диаграммы процесса

Первый этап отличается особой важностью в рассматриваемом процессе, поскольку формулирует и фиксирует четкие критерии для конечного желаемого результата. На этом этапе составляется сценарий использования модели и формулируется бизнес-цель разработки. Уточняется, какой тип скоринга будет охватывать модель, в каких сегментах и временных интервалах она должна функционировать, какую аналитическую задачу –

классификации, регрессии, кластеризации или иную – предстоит решить. Определяется, кто будет основным пользователем модели, как именно её результаты будут интерпретироваться в рамках бизнес-процессов и какие эффекты ожидаются от её применения.

Далее проводится анализ ограничений среды разработки и внедрения. Учитываются характеристики доступных технических ресурсов, перечень применяемых инструментов и возможности эксплуатации модели в производственной среде. На этом же этапе формулируются требования к качеству модели и закладываются принципы её архитектурной реализации. Особое значение имеет выбор целевой переменной, то есть параметра, который модель будет предсказывать. В рамках описания требований также актуально определение временного горизонта для наблюдений, критериев включения и невключения данных в обучающую выборку, определение предварительного набора признаков, потенциально зависящих от таргета. Эти действия формируют информационную основу и функциональные контуры будущей модели и позволяют перейти к выбору методов построения, распределению ролей в команде проекта и составлению регламента по срокам выполнения. Таким образом, утверждается проект новой модели и сценарий её использования, регламентирующий последующее внедрение. Приведенное описание представлено декомпозицией подпроцесса на Рисунке 3.

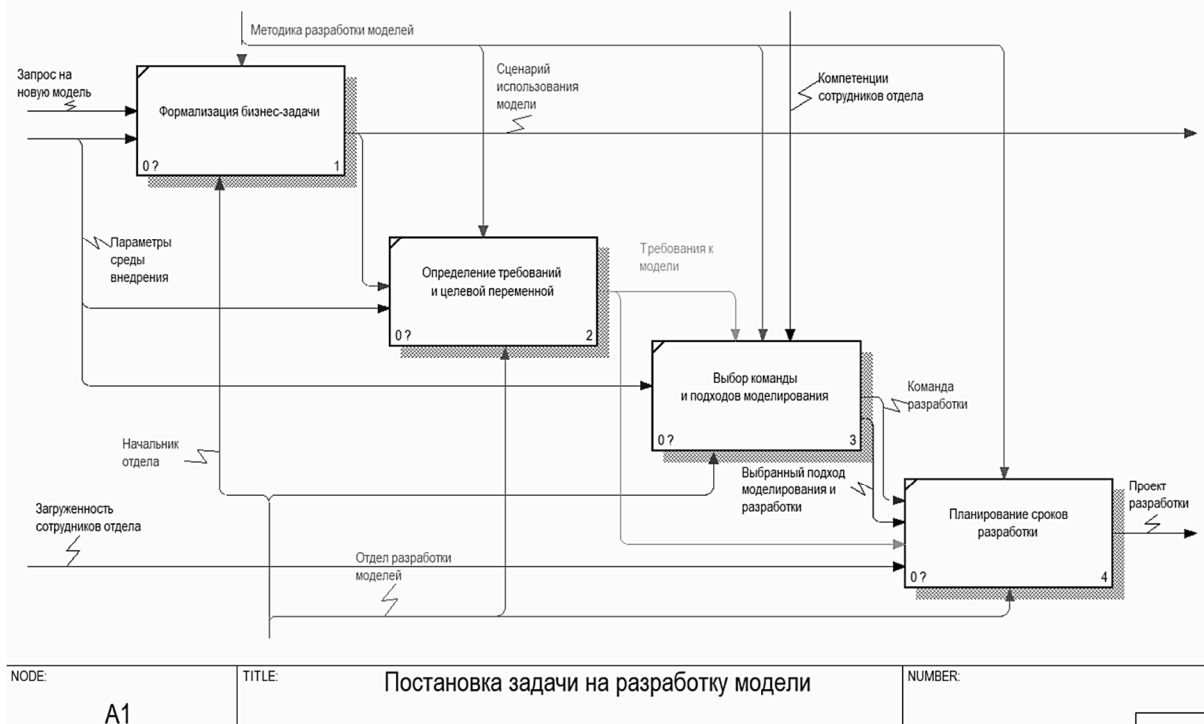


Рисунок 3 – Декомпозиция подпроцесса «Постановка задачи на разработку модели»

Второй этап декомпозиции контекстной диаграммы фокусируется на поиске и выборе источников данных, содержащих актуальную и достаточную для обучения информацию. Источниками могут выступать витрины с анкетными и историческими данными клиентов, хранилища транзакций по банковским картам, отчёты о характеристиках кредитного портфеля в разрезе отдельных контрактов, а также графики платежей, включая сведения о просроченных платежах. Дополнительно могут использоваться данные о действиях по взысканию задолженностей, фактах банкротства и реструктуризации долгов. Все отобранные источники и поля в дальнейшем находят отражение в процедурах сбора данных, которые должны соответствовать ранее зафиксированным требованиям к выборке.

После сбора данных начинается этап их анализа и предобработки, в рамках которого формируются дополнительные признаки, направленные на повышение информативности

выборки. Такие признаки могут отражать поведенческие аспекты клиентов, например, в виде динамических коэффициентов и долей. Проводится очистка от шумов и выбросов, нормализация значений и структурирование данных для последующего обучения модели. Особое внимание уделяется формированию обучающей, тестовой и отложенной выборок. Обучающая выборка предназначена для построения модели и выявления закономерностей в данных, тестовая выборка позволяет оценить точность модели на ранее не использованных данных, что даёт представление о её обобщающей способности. Отложенная (out-of-time) выборка формируется на более поздних временных интервалах и предназначена для имитации условий реального применения модели после внедрения, обеспечивая оценку её устойчивости и стабильности во времени.

Следующий шаг – построение базовой (baseline) модели, за счет которой возможно зафиксировать «исходный» уровень качества, принимаемый за точку отсчета для последующих улучшений. Перед построением из полного набора переменных выбираются те, которые демонстрируют наибольшую прогностическую силу и связь с целевой переменной, что может оцениваться как с помощью статистических методов, так и с учётом экспертных оценок. Отобранные переменные могут быть дополнительно обработаны – закодированы, сгруппированы в бины или масштабированы, – и составляют конечный набор полей для сформированных выборок. На основе обработанных данных обучается базовая модель без сложной настройки параметров. Её цель – получить начальные метрики качества (точность, стабильность, предсказательная сила) на тестовой и отложенной выборках и определить направления оптимизации. Переменные с признаками нестабильности или переобучения исключаются, модель пересобирается до достижения устойчивого результата. Этап подробно рассмотрен как декомпозиция соответствующего подпроцесса на Рисунке 4.

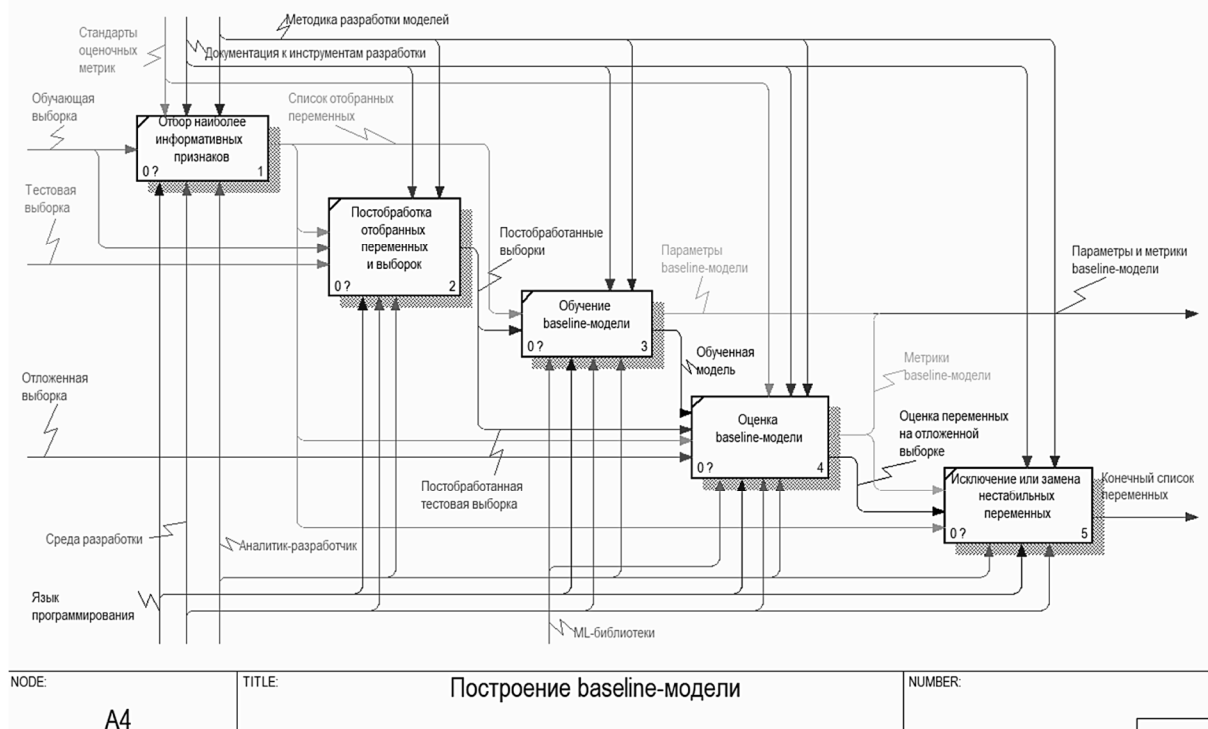


Рисунок 4 – Декомпозиция подпроцесса «Построение baseline-модели»

С использованием конечного набора признаков обучается основная модель до момента нахождения оптимальных параметров, что также сопровождается оценками метрик и стабильности на различных временных интервалах. Ключевым отличием от предыдущего этапа является наличие калибровки итоговой модели. Калибровка необходима для того, чтобы прогнозы модели лучше соответствовали реальному поведению клиентов. Несмотря на то, что модель может хорошо различать классы, её

предсказания могут быть смещёнными, что систематически будет завышать или занижать вероятность дефолта клиентов. Калибровка позволяет "подстроить" вероятности так, чтобы прогнозируемые значения совпадали с фактической частотой событий и улучшали интерпретируемость модели [9]. На практике калибровка особенно важна для моделей, используемых в реальных продуктах, где точность вероятностей напрямую влияет на решения и финансовые результаты. Финальная модель должна демонстрировать устойчивое качество и быть пригодной для бизнес-интерпретации.

После завершения разработки модель сопровождается отчётом, содержащим описание источников данных, финальных признаков, использованных методов и достигнутых метрик качества. Модель проходит валидацию – внутреннюю или внешнюю – где проверяется соответствие результатов проекта нормативным требованиям, корректность расчётов и прохождение статистических тестов, таких как биномиальный тест. При условии успешного прохождения всех проверок модель внедряется в производственную среду. Это может быть реализация в виде модуля на стороне банковской системы, загрузка модели во внешний сервис или интеграция во внутренние процессы принятия решений, в зависимости от того, что было определено в сценарии использования.

Заклучение. Таким образом, разработка моделей кредитного скоринга представляет собой структурированный и многоэтапный процесс, охватывающий как технические аспекты построения, так и проработку бизнес-целей и требований регуляторов. Четкая последовательность этапов обеспечивает оперативную и высококачественную разработку устойчивых во времени моделей, которые, в свою очередь, автоматизируют деятельность кредитного скоринга. Результаты работы должны способствовать повышению точности и эффективности оценки кредитоспособности клиентов банка, а также улучшению процессов управления кредитными рисками в банковском бизнесе.

Список литературы:

1. Политика управления рисками Банка России // Банк России. М., 2020. / URL: <https://cbr.ru/Content/Document/File/36486/policy.pdf> (дата обращения: 11.05.2025).
2. Addo, P.M., Guegan, D., Hassani, B. Credit risk analysis using machine and deep learning models // Risks. MDPI, 2018. / URL: <https://doi.org/10.3390/risks6020038> (дата обращения: 11.05.2025).
3. Как мы внедрили ML в приложение с почти 50 миллионами пользователей. Опыт Сбера // Блог компании Сбер. 2019. / [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/companies/sberbank/articles/461747/> (дата обращения 11.05.2025).
4. «Тинькофф» начал автоматически отклонять заявки мошенников на кредиты с помощью ИИ / [Электронный ресурс]. – Режим доступа: URL: <https://rb.ru/news/tinkoff-ai/> (дата обращения 11.05.2025).
5. Как мы строили самую большую модель кредитного скоринга в сегменте МСБ // Блог компании Точка, 2022. / [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/companies/tochka/articles/696226/> (дата обращения 11.05.2025).
6. Смирнов В. Нейросетевой подход к кредитному скорингу на данных кредитных историй // Блог компании Альфа-Банк, 2022. / [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/companies/alfa/articles/680346/> (дата обращения 11.05.2025).

7. Как избавить датасаентистов от рутины: опыт автоматизации моделирования в Альфа-Банке / [Электронный ресурс]. – Режим доступа: <https://futurebanking.ru/post/4089> (дата обращения 11.05.2025).
8. Кириллина, Ю. В. Управление бизнес-процессами : учебное пособие / [Электронный ресурс]. – М.: РТУ МИРЭА, 2024. – Электрон. опт. диск (ISO).
9. Niculescu-mizil, A., Caruana, R. Predicting good probabilities with supervised learning // Machine Learning, Proceedings of the Twenty-Second International Conference. Bonn, Germany, 2005.

References:

1. Bank of Russia Risk Management Policy // Bank of Russia. Moscow, 2020. / URL: <https://cbr.ru/Content/Document/File/36486/policy.pdf> (accessed 11.05.2025).
2. Addo, P.M., Guegan, D., Hassani, B. Credit risk analysis using machine and deep learning models // Risks. MDPI, 2018. / URL: <https://doi.org/10.3390/risks6020038> (accessed 11.05.2025).
3. How We Integrated ML into an App with Nearly 50 Million Users. Sberbank's Experience // Sberbank Blog, 2019. / [Electronic resource]. – Available at: <https://habr.com/ru/companies/sberbank/articles/461747/> (accessed 11.05.2025).
4. Tinkoff Begins Automatically Rejecting Fraudulent Loan Applications Using AI / [Electronic resource]. – Available at: <https://rb.ru/news/tinkoff-ai/> (accessed 11.05.2025).
5. How We Built the Largest Credit Scoring Model in the SME Segment // Tochka Bank Blog, 2022. / [Electronic resource]. – Available at: <https://habr.com/ru/companies/tochka/articles/696226/> (accessed 11.05.2025).
6. Smirnov, V. Neural Network Approach to Credit Scoring Based on Credit History Data // Alfa-Bank Blog, 2022. / [Electronic resource]. – Available at: <https://habr.com/ru/companies/alfa/articles/680346/> (accessed 11.05.2025).
7. How to Free Data Scientists from Routine: Automation Experience at Alfa-Bank / [Electronic resource]. – Available at: <https://futurebanking.ru/post/4089> (accessed 11.05.2025).
8. Kirillina, Y.V. Business Process Management: A Study Guide / [Electronic Resource]. – Moscow, RTU MIREA, 2024. – CD-ROM (ISO).
9. Niculescu-mizil, A., Caruana, R. Predicting good probabilities with supervised learning // Machine Learning, Proceedings of the Twenty-Second International Conference. Bonn, Germany, 2005.