

Лебедев Владимир Александрович

студент

Фомичева Татьяна Леонидовна

канд. экон. наук, доцент

ФГБОУ ВО «Финансовый университет
при Правительстве Российской Федерации»

г. Москва

РОЛЬ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В КРЕДИТНОМ СКОРИНГЕ

***Аннотация:** в статье рассматривается сущность кредитного скоринга и роль искусственного интеллекта в кредитном скоринге.*

***Ключевые слова:** кредитный скоринг, искусственный интеллект, оценка кредитоспособности заемщика, математическая модель.*

Кредитный скоринг – оценка кредитоспособности заёмщика (кредитных рисков), основанная на численных статистических методах – известный кейс применения математических методов в банковской сфере.

Темпы, с которыми организации вкладывают средства в искусственный интеллект (ИИ), продолжают стремительно расти. Исследования консалтинговой компании «Accenture» показывают, что финансовые организации, которые инвестируют в ИИ и робототехнику, могут увеличить свои доходы более чем на треть (34%) к 2022 году. Кредитный скоринг – одна из областей, в которой банковская индустрия использует ИИ. Технология извлекает идеи из данных о контрактах, прошлых расходах и будущих доходах, чтобы обеспечить ряд кредитных решений. Но использование ИИ для кредитного скоринга указывает на некоторые риски, как утверждает Ральф Олхаузен, директор по развитию бизнеса в финансовом учреждении PPRO Group.

«Если алгоритмы не будут тщательно протестированы, чтобы избежать повторения смещения, которое уже существует в данных, или введения новых смещений путем создания ложных корреляций, то существует опасность того, что

АИС кредитного рейтинга может несправедливо отказать некоторым людям в доступе к финансовым услугам», – отмечает он.

Использование технологий Data Mining в банковском секторе

Существование банка зависит от оперативности принятия стратегических решений. Большая часть того, что нужно для принятия стратегического решения уже хранится в банках в виде огромных объемов данных, которые собираются ежедневно для решения основных задач во всех сферах банковского бизнеса. Использование технологий «Big Data» дает банкам преимущества перед своими конкурентами. Оперативное извлечение полезной информации из больших архивов данных и ее анализ в разных аспектах повышает эффективность работы банка в целом и помогает лучше понять поведение клиентов. Благодаря технологиям «Big Data» системы безопасности умеют оперативно анализировать колоссальные массивы информации самого разного происхождения и эффективно бороться с мошенничеством. Технологии «Big Data» позволяют банку использовать полу структурированные данные, чтобы выявить скрытые связи между клиентами для определения их уровня кредитоспособности. В целом технологии «Big Data» благодаря анализу клиентских баз, а также неструктурированной информации из социальных сетей и других различных источников позволяет разрабатывать более сфокусированные маркетинговые предложения, больше соответствовать ожиданиям разных целевых аудиторий, лучше оказывать влияние на потребительское поведение клиентов. Благодаря технологиям «Big Data» банки получают возможность добавить в свою деятельность качественно новую аналитику и элементы искусственного интеллекта. Информация о клиентах, поставщиках, операциях компании и многом другом распределена между десятками операционных систем и систем обработки транзакций, которые находятся в разных бизнес-подразделениях. Появляется необходимость в объединении всей информации вместе с целью получить истинные знания. В этом помогает технология «Data Mining», которая является частью профилирования заёмщиков в банковской сфере. Инновационные технологии «Data Mining» позволяют обрабатывать информацию с целью автоматического поиска шаблонов, которые

характеризуют какие-либо фрагменты неоднородных многомерных данных. Область применения «Data Mining» не имеет границ – она присутствует везде, где имеются какие-либо данные. Методами «Data Mining» сегодня заинтересованы, в первую очередь, банки. Профилирование наилучших достижений, как один из методов «Data Mining», позволяет выявить основные характеристики наиболее успешных регионов, филиалов, клиентов для планирования дальнейшей деятельности банка. Профилирование также используется в банках для оценки риска при выдаче кредитов.

Обратим внимание на некоторые задачи технологии «Data Mining»:

1. Определение возможной некредитоспособности клиента банка или так называемая задача «Выдавать ли кредит клиенту?»
2. Выявление мошенничества.
3. Сегментация клиентов.



Рис. 1

Использование технологий Machine Learning в банковском секторе

Используя имеющиеся данные, розничные банки на разных уровнях управления решают задачи:

- 1) управления рисками;
- 2) клиентоориентированности;
- 3) оптимизации операционной деятельности.

Перед современным банком по каждой из перечисленных задач стоит целый ряд проблем. Опишем их тезисно ниже.

1. Модели расчета банковских рисков, скоринговые модели все более и более усложняются, одновременно со сложностью моделей (качественной сложностью) растет и объем анализируемых данных (количественная сложность). А все увеличивающееся разнообразие способов оплаты (mPOS, мобильные платежи) и количество получателей платежей открывает все новые и новые фронты борьбы с мошенничеством.

2. Клиенты банков становятся все более и более требовательными как к банковскому сервису, так и к банковским продуктам. Клиентам уже не нужна «просто пластиковая карта». Клиентам нужен составной продукт, который позволит им покупать товары (offline и online); продукт, с постоянным и удобным доступом (со смартфона, планшета, домашнего ПК); продукт, который позволит экономить средства (скидки по карте, cashback, бонусы).

3. Увеличение конкуренции, как со стороны виртуальных банков, так и со стороны стартапов, предоставляющих финансовые сервисы (P2P-кредитование, агрегаторы банковских счетов), вынуждают игроков на рынке банковских услуг снижать маржинальность многих банковских продуктов. В такой ситуации «полноценным» банкам важно уменьшать операционные издержки.

Глобальной же проблемой является тот факт, что все возрастающие объемы как внутренних данных банка, так и внешних данных требуют от банка качественно новых подходов к хранению и обработке данных.

Рассмотрим, как симбиоз концепций Больших Данных и методов машинного обучения позволяет эти проблемы решать:

– сегментация клиентов – первая ступень на пути решения большинства обсуждаемых ниже проблем. Цель сегментации – деление клиентов на группы (кластеры) со схожим поведением. То, к какой группе будет отнесен клиент, зависит от совокупности аспектов поведения, по которым проходит сегментация.

Сама по себе сегментация не дает никаких ответов (и тем более точных), она лишь показывает наличие/отсутствие паттернов или определенной структуры в

данных. Источники данных, по которым будет проводиться сегментация, зависят от конкретной задачи, для которой идет поиск паттерна. В большинстве случаев это банковская анкета клиента и лог транзакций – журнал финансовых операций, проведенных всеми клиентами банка за определенный период времени.

Методы машинного обучения, которые применяются для решения данной задачи – это методы кластерного анализа. Метод k-средних (k-means) в большинстве случаев отлично справляется с задачей кластеризации по клиентам банка.

Одной из интересных идей сегментации клиентов является идея группировки клиентов на основе результатов семантического анализа общения с клиентами. Для этой задачи источниками данных являются записи разговоров с клиентом в call-центре, история переписки со службой поддержки банка, анализ публичных постов клиента, оставленных в социальных сетях.

Сегментация клиентов относится к задаче пакетной обработки данных (batch processing) и, в общем случае, не является задачей реального времени (real-time processing);

– обнаружение мошенничества. Можно выделить следующие способы фрода (от англ. Fraud – мошенничество), приводящие к несанкционированному снятию/пополнению денежными средствами банковского счета клиента:

- 1) операции с украденными пластиковыми картами;
- 2) операции с украденными идентификационными данными клиента, карты (в том числе скимминг);
- 3) мошенничество при оплате пластиковой картой через интернет-магазины;
- 4) использование украденных учетных данных интернет- и мобильных банков (в том числе фишинг), взлом и эксплуатация уязвимостей мобильных и интернет-банков;
- 5) инсайд (мошенничество со стороны сотрудников банка).

Для решения задач [1–3] необходимо провести первичную сегментацию клиентов, определить поведенческий паттерн движения денежных средств для конкретного клиента. Далее имеет смысл использовать такие методы как Дерево

решений (Decision Tree) или нейросети (Neural Network), чтобы определить вероятность того, что платеж является мошенническим.

Источники данных: логи транзакций; личные сведения о клиенте, известные банку (такие как адрес проживания). Для задачи [3] хорошей практикой будет использовать сведения от мерчанта (IP адрес, с которого пришел запрос на проведение платежа, адрес доставки товара, имя владельца учетной записи), справочники адресов (определение приблизительного по индексу, по IP адресу).

Для решения задачи [4] необходимо анализировать лог интернет-банка/мобильного приложения и логи backend-сервисов, работающих с этими приложениями, а затем с помощью методов обнаружения аномалий (Anomaly Detection) искать нехарактерную активность (слишком частые попытки авторизации, слишком высокая частота внешних денежных переводов) по подмножеству учетных записей.

Задача [3] – выявление мошенничества при online-оплате посредством пластиковых карт – одна из наиболее интересных историй использования data science в банковском секторе. Так известно, что международная платежная система Visa в 2011 году разработала новую версию антифрод-системы с использованием методов Больших Данных. Заявлено, что за год система останавливает мошеннические действия на сумму ~ \$2 млрд. Компания IBM в своем отчете заявила, что использование методов машинного обучения в антифрод-системе позволило «отлавливать» на 15% больше мошеннических платежей и на 50% уменьшить ложное срабатывание антифрод-системы.

Еще одна интересная идея по улучшению работы подобных систем – использования геолокационных данных о нахождении клиента, предоставляемых сотовым оператором. С развитием технологий Apple Pay, Android Pay, Samsung Pay учет информации о географическом нахождении клиента для принятия решения о том, является ли платеж мошенническим, кажется вполне логичным элементом системы безопасности проведения финансовых транзакций.

Задачи выявления мошенничества – time sensitive. Так задачи [1–3] являются near realtime задачами, где большие задержки в обработке данных – это прямые и косвенные (репутационные) убытки для банка.

«Big Data + Machine Learning» помогает решать следующие банковские проблемы на качественно другом уровне:

- сегментация клиентов;
- обнаружение мошенничества;
- кредитный скоринг.

Для решения перечисленных выше задач современный уровень развития Data Science позволяет без участия аналитика:

- выделять из данных имеющуюся структуру и закономерности (кластерный анализ, регрессионный анализ);
- рассчитывать вероятность происхождения некоторого события в будущем (предиктивный анализ);
- выделять необычное поведение (обнаружение аномалий);
- определять эмоциональную окраску сообщения, выделять из него субъекта и объекта (латентно-семантический анализ).

Принимая во внимание стремительно растущие объемы данных, современные же Big Data подходы позволяют:

- распределено запускать в вычислительном кластере задачи, выполняющие запросы к терабайтам (и даже петабайтам) данных;
- масштабировать выполнение запросов на все увеличивающихся объемах данных.

Типичная последовательность действий специалиста при поиске закономерностей в наборе данных с использованием алгоритмов обучения с учителем изображена на иллюстрации ниже.



Рис. 2

Заключение

На основании проведенного в рамках статьи исследования можно сделать вывод, что искусственный интеллект поможет банкам существенно автоматизировать процессы и повысить качество обслуживания клиентов. Естественно, полностью заменить человека не удастся, и в ближайшем будущем не стоит ожидать банковских систем, которые полностью находятся под управлением ИИ. Тем не менее искусственный интеллект способен автоматизировать однотипные банковские процессы и улучшить обслуживание клиентов. В настоящее время

застой в экономике отрицательно сказывается на рынке кредитования: объем займов снижается, а их качество падает – растет число просроченных выплат и невозвратов. Особую актуальность в таких условиях получает ПО для скоринга, благодаря которому производится оценка платежеспособности потребителей. Такой инструмент позволяет банку найти оптимальный баланс между количеством выданных займов и процентом «плохих» кредитов. Каждый банк определяет приемлемый процент невозврата кредитов в индивидуальном порядке в зависимости от проводимой политики, включая процентную ставку, стоимость привлечения средств, типы кредитных продуктов и т. п. С технической точки зрения существует три модели реализации скоринга.

Первый вариант – это применение самописного программного обеспечения. Такой вариант является наиболее затратным, так как индивидуальная разработка подразумевает значительные вложения не только в лицензии на программное обеспечение для построения скоринговых карт, но и экспертизу, аналитику. Цена ПО ведущих поставщиков, как на российском, так и на мировых рынках варьируется от сотен тысяч до миллионов рублей. Помимо этого, необходим значительный штат аналитиков, который будет заниматься разработкой и поддержкой скоринговых моделей кредитной организации.

Второй вариант применения скоринга – это покупка и интегрирование готового продукта стороннего разработчика. Это помогает сэкономить на лицензиях на ПО и не требует внушительных ресурсов для анализа и поддержания скоринга. Но приобретаемая модель учитывает всего лишь «среднюю температуру» по рынку, поэтому, как правило, после внедрения банк нанимает собственный штат аналитиков и корректирует модель под себя.

Третий способ – это использование внешнего сервиса, который предоставляется кредитным бюро. Банк отправляет запрос в БКИ и в ответ получает оценку кредитоспособности клиента в баллах. В этом случае тарифицируется каждое обращение, стоимость одного обращения находится в пределах 15 руб. В данном случае банк не нуждается в найме собственных аналитиков, но

приходится довольствоваться общей моделью без учета особенностей банка. Такой вариант идеально подходит небольшим организациям.

Список литературы

1. Фомичева Т.Л. Применение Datamining и Machinelearning в процессе кредитного скоринга / Т.Л. Фомичева, С.В. Савина, Е.А. Викулина // Евразийский юридический журнал. – 2018. – №7 (122). – С. 413–414 [Электронный ресурс]. – Режим доступа: <https://elibrary.ru/item.asp?id=35351595>
2. Аглицкий И.С. Кредитный скоринг – бремя или благо? // Экономика и управление: проблемы, решения. – 2017. – Т. 7. – №8. – С. 114–115 [Электронный ресурс]. – Режим доступа: <https://elibrary.ru/item.asp?id=30071684>
3. Бандалиева А.М.К. Кредитный скоринг на основе искусственного интеллекта / А.М.К. Бандалиева, Н.В. Давудян, О.Г. Акопян, Н.И. Ломакин // Актуальные проблемы развития хозяйствующих субъектов, территорий и систем регионального и муниципального управления: Материалы XIII международной научно-практической конференции / Под ред. Ю.В. Вертаковой. – 2018. – С. 24–27 [Электронный ресурс]. – Режим доступа: <https://elibrary.ru/item.asp?id=35117817>
4. Маряшин Р.В. Кредитный скоринг как инструмент эффективной оценки кредитоспособности // 2018. – №3 (105). – С. 33–35 [Электронный ресурс]. – Режим доступа: <https://elibrary.ru/item.asp?id=35292080>