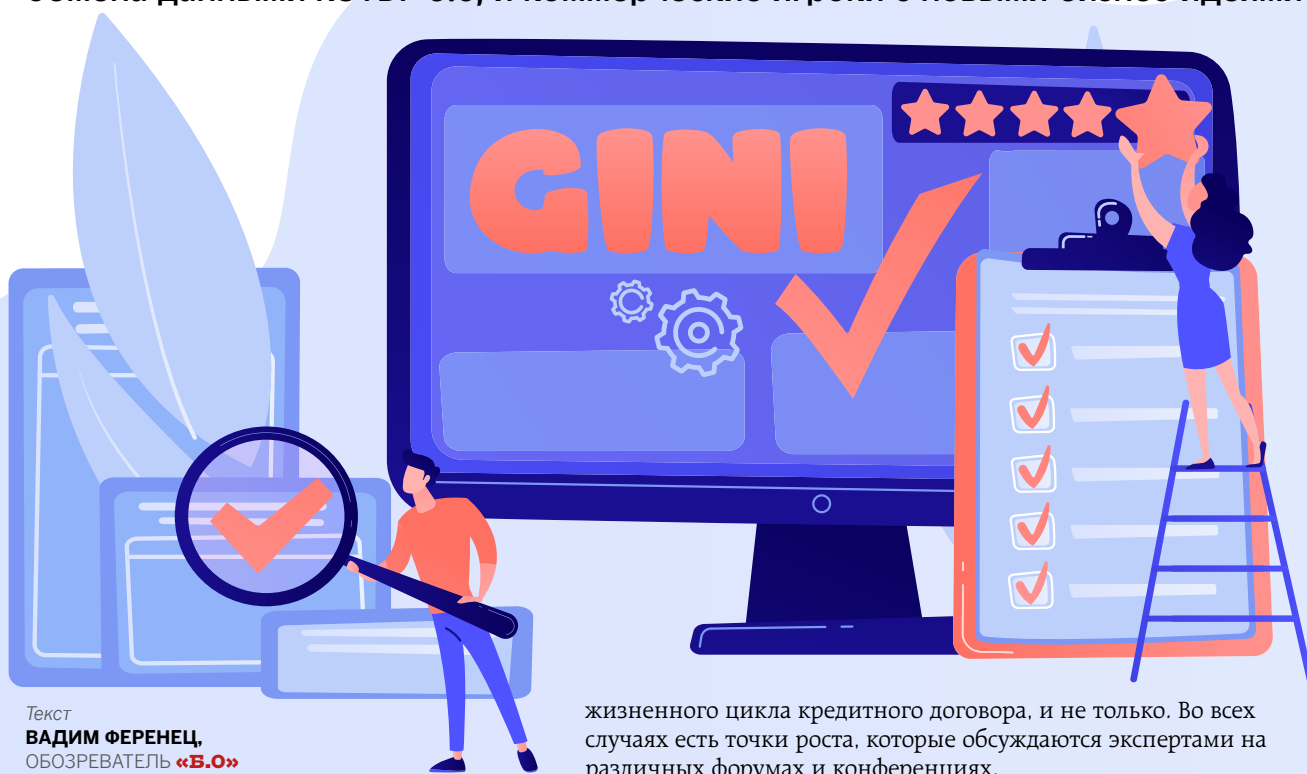


Что нам стоит поднять Gini?

Работа по повышению предсказательной силы (Gini) скорингов ведется постоянно. В этом участвуют и Банк России, который запустил формат обмена данными RUTDF 5.0, и коммерческие игроки с новыми бизнес-идеями



Текст
ВАДИМ ФЕРЕНЦ,
 ОБОЗРЕВАТЕЛЬ «Б.О.»

В последнее время эксперты, связанные с кредитным скорингом, все чаще стали говорить о том, что в условиях, когда у крупных банков сложился практически идентичный пул поставщиков данных, весьма вероятно перспектива получения всеми схожих скоринговых баллов. А это означает, что наступает эпоха конкуренции моделей. Под моделями здесь понимаются как математические алгоритмы, которые рассчитывают собственно скоринговый балл, так и экономические модели ведения бизнеса, которые, естественно, должны приносить доход от кредитного процесса с учетом всего

жизненного цикла кредитного договора, и не только. Во всех случаях есть точки роста, которые обсуждаются экспертами на различных форумах и конференциях.

Наиболее содержательными в этой проблематике стали 7-й Ежегодный форум скоринговых технологий (ScorFest), прошедший в июле 2023 года, и Форум Scoring Day 12, состоявшийся в сентябре этого же года.

Экономика скоринга

ScorFest выделился своей экспертной дискуссией на тему «Сколько стоит один Gini?» в контексте обсуждения затрат на повышение предсказательной/ранжирующей силы скоринга на один пункт на практике, а также выяснения, какая стоимость обработки одной заявки экономически оправдана в сложившихся условиях.

Что касается типов моделей, то общее мнение экспертов таково: логистическая регрессия и градиентный бустинг — сферы ИИ и ML, которые лучше всего работают с тем главным, что есть у любого банка: данными по транзакциям клиентов.

Из чего складывается себестоимость скоринга? **Антон Исправников**, заместитель начальника управления моделирования КИБ

и СМБ департамента анализа данных и моделирования банка ВТБ, привел «среднюю по больнице» пропорцию: «70% затрат приходится на IT-инфраструктуру, а на все остальное — 30%».

Говорить про снижение затрат на IT в рамках конкретного банка крайне сложно, учитывая происходящие процессы импортозамещения и дефицит кадров. А вот сегмент «все остальное» имеет гораздо большие вариативные границы (например, за счет экономии на повсеместном использовании ИИ, а также автоматизации внедрения моделей в промышленную эксплуатацию с помощью AutoML).

Еще один вариант снижения затрат заключается в использовании результатов скоринга в иных процессах, например продажах: можно повысить продажи, не наращивая при этом уровня риска выше заданных лимитов. Этот эффект приводит, по словам **Дмитрия Рузанова**, Chief Data Scientist Альфа-Банка, к дополнительной монетизации имеющихся данных, что входит как элемент в стратегию данного банка.

Алексей Волков, директор по маркетингу Национального бюро кредитных историй (НБКИ), добавил: «Работа по повышению предсказательной силы (Gini) скорингов ведется постоянно, в том числе за счет известных, а также новых алгоритмов искусственного интеллекта. Это актуально для тех моделей, где разделяющая способность более важна, чем интерпретируемость».

Очевидно, имеет ценность не только сказанное экспертами, но и обсуждаемое лишь вскользь, а именно деятельность Банка России по усилению инфраструктурной компоненты всей финансовой отрасли. В частности, с 1 апреля 2023 года вступило в силу Указание Банка России № 6300-У от 24 октября 2022 года, которое внесло изменения в Положение Банка России № 758-П «О порядке формирования кредитной истории».

Регулятор установил, что с 1 октября 2023 года взаимодействие между источниками формирования кредитных историй и самими бюро кредитных историй будет происходить исключительно с применением формата обмена данными RUTDF 5.0, предлагающего расширенные возможности по сбору информации о заемщиках и анализу их платежеспособности. RUTDF 5.0 за счет большей детализации позволяет точнее оценить достоверность передаваемой информации, что повышает качество скоринга, увеличивая бизнес-показатели и стоимость процедуры.

Алексей Волков рассказал: «Мы видим отличную возможность по увеличению качества скоринга и показателя Gini после введения нового формата. Более того, мы уже этим занимаемся и достигли определенных успехов в соответствующей разработке. Отличие от предыдущих форматов заключается в расширении и большей детализации данных. Сам формат кредитной истории RUTDF версии 5.0 и выше основан на Положении Банка России 758-П «О порядке формирования кредитной истории» и в настоящее время доступен всем пользователям и источникам формирования кредитных историй: банкам, МФО, МКК и т.д.».

«Карма», меняющая мир

На форуме Scoring Day можно было узнать подробности об эффектах совместного использования и обработки данных в безопасном формате различных поставщиков в интересах индустрии скоринга, которые базируются на инфраструктуре коммерческих игроков. Характерно, что в этой части дискуссии речь шла как о традиционных нейронных сетях, так и о генеративно-состязательных сетях (Generative adversarial network, GAN), а также о больших языковых моделях (Large language models, LLM).

Связано это было с изменением моделей ведения бизнеса банками и интеграции их финансовых сервисов в большие экоси-

стемы вокруг крупных игроков из телекома, ретейла и интернет-гигантов. Очевидно, что скоро с банком можно будет общаться голосом и выяснить, что именно послужило причиной отказа в получении кредита или займа и как это можно исправить.

Из выступления **Никиты Зелинского**, Chief Data Scientist компании МТС, можно было сделать несколько выводов. Во-первых, МТС активно включилась в совместную с банком ВТБ работу в области Data Fusion и безопасного совместного объединения данных с помощью технологии криптоанклавов. МТС и на своих, и на объединенных данных участников криптоанклава обучает нейросетевые модели. Это один из эффективных путей повышения Gini при уменьшении затрат на IT-инфраструктуру.

Второе, на чем фокусируется МТС как один из крупнейших поставщиков готовых скоринговых баллов, это погружение в эмбединги (embedding) — процесс преобразования слов или текста в набор чисел, числовой вектор. Векторы можно сравнивать между собой, чтобы определить, насколько два текста или слова похожи по смыслу. Есть заметные успехи с эмбедингами с графовых нейросетей. В ближайших планах — создание текстовых эмбедингов для автоматической выдачи мотивированного суждения, например, Росфинмониторингу о том, почему конкретному клиенту был присвоен именно такой скоринговый балл. По сути, это революция в раскрытии данных о конкретном скоринге.

За скромными словами Никиты Зелинского о достижениях явно просматривается недосказанное по поводу активизации усилий в области экосистем и финансовых платформ для предложения физическим лицам услуг по подбору потребительских кредитов.

Отметим, что успехи американского финтех-единогорога Credit Karma — американской платформы для управления кредитными и финансовыми операциями, а также для налоговой подготовки, мониторинга не востребуемых баз данных и инструмента для выявления ошибок в области кредитного отчета — были отмечены множеством спикеров. Если эта финтех-компания смогла изменить скоринговую инфраструктуру в США, то рано или поздно кто-то сможет сделать это и в нашей стране.

Но пока это будущее не наступило, **Александр Чернощекин**, старший вице-президент Промсвязьбанка, рассказал, как с помощью глобального скоринга юриц в области МСП расширить кредитный портфель *в десять раз* без роста дефолтности и себестоимости продукта. Так что и у нас в стране есть эксперты, способные ответить на вопрос: «Что нам стоит поднять Gini?».