Сергей АФАНАСЬЕВ

Есть два ключевых драйвера развития антифрода: с одной стороны, мошенники быстро адаптируют свои схемы, обходя банковскую защиту, из-за чего та устаревает; с другой стороны — новые банковские продукты и технологии порождают и новые уязвимости. В результате банки вынуждены реагировать на новые вызовы с удвоенной скоростью. Какова роль аналитики в банковском антифроде? Как при помощи аналитических подходов увеличить точность оценки вероятности мошенничества POS-партнеров и повысить эффективность биометрической системы?

Как с помощью аналитики повысить эффективность банковского антифрода?

Сергей АФАНАСЬЕВ, *КБ «Ренессанс Кредит» (ООО), начальник управления расследования мошенничества*

Банковский антифрод является закрытой сферой. При этом инструментарий в антифроде, с одной стороны, быстро устаревает, а с другой — так же быстро пополняется. Именно поэтому межотраслевые стандарты в антифроде существуют скорее для описания базовых принципов противодействия мошенничеству, а не для стандартизации актуального инструментария. По этим причинам основные идеи для разработки антифрод-инструментов мы (антифродеры) черпаем из трех дисциплин (рис. 1):

1) аудит — дает различные инструменты проверок, ревизий, инспектирований, контроля и т.д.;

2) криминалистика/криминология[1] — дает инструментарий для детектирования мошенничества и проведения расследований;

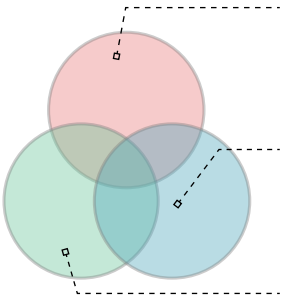
3) аналитика — позволяет использовать математический аппарат для разработки антифрод-инструментов и их оценки.

[1] Традиционно считается, что криминалистика и криминология — разные науки. Тем не менее, сложно отрицать, что криминология влияет на криминалистику и наоборот. Например, понимая психологию мошенника, легче придумать инструменты для его поимки.

Как с помощью аналитики повысить эффективность банковского антифрода?

Рисунок 1

Три составляющие банковского антифрода



Аудит
Сплошная и выборочная проверки, документальная ревизия, аналитическая проверка, инспектирование и т.д.

Криминалистика
Сбор улик, задержания, допросы, очные ставки, обыски, криминалистическая экспертиза и пр.

Аналитика
Поиск аномалий, выявление fraud-концентраций, разработка правил и моделей, финансовая оценка и др.

Если говорить о функциональном разделении, то за аудиторский инструментарий обычно отвечает подразделение внутреннего аудита, за криминалистику — служба безопасности, за аналитику — подразделение риск-менеджмента.

По типам контроля антифрод-инструменты делятся на три вида[1]:

1) директивные (directive) — это различные инструкции, обучающие материалы, договоры, мотивационные схемы и т.п., в которых описывается, что запрещено, а что разрешено;

2) детективные (detective) — инструменты, с помощью которых выявляют мошенничество;

3) превентивные (preventive) — инструменты, которые направлены на предотвращение мошеннических действий.

Такая типизация позволяет достаточно просто и наглядно понять, какие инструменты нужны в бизнес-процессах банка, а какие нет. Например, если в банке разрабатывается новый бизнес-процесс, то антифрод-специалисты оценивают:

— какие директивные инструменты необходимы для этого процесса, чтобы довести до сотрудников, партнеров или клиентов информацию о разрешенных и запрещенных действиях с их стороны;

— какие детективные мероприятия необходимо реализовать, чтобы выявлять мошеннические действия и нарушения инструкций;

[1] В литературе по риск-менеджменту можно встретить другие виды типизаций инструментов, например: превентивные, детективные и корректирующие (corrective); или превентивные, детективные и компенсирующие (reactive) и т.д. А в стандартах по информационной безопасности (CISSP) используется более детализированная сегментация: preventative, deterrent, detective, corrective, recovery, compensation, directive, administrative, logical/technical, physical.

Сергей АФАНАСЬЕВ

— какие есть уязвимости в процессе, какими превентивными инструментами и мероприятиями можно закрыть эти уязвимости.

Обычно директивные инструменты — самые дешевые для банка и хорошо действуют на сознательных сотрудников, партнеров или клиентов. Детективные инструменты дороже и, как правило, требуют штата сотрудников службы безопасности и антифрод-специалистов. При этом детективные инструменты хорошо дополняют директивные (если сотрудник нарушил инструкции, необходимо это выявить). Ну и, конечно, главная цель антифрода — предотвратить мошенничество, поэтому превентивные инструменты можно считать самыми эффективными, но и часто самыми дорогими[1].

Помимо сегментации инструментов по типам контроля (типизация снизу вверх), мы в банке используем «карту антифрод-аудита». Согласно этой карте все известные нам антифрод-инструменты и мероприятия (не обязательно внедренные в банке) разбиты на три блока, каждый из которых состоит из нескольких тематик (рис. 2):

Рисунок 2

Карта антифрод-аудита (пример)

Бизнес-процессы		Аналитика		Антифрод-системы	
Персонал	Проверка сотрудников	Репорты	Fraud Report	Внутренние системы	Warning System
	Обучение		Concentration Report		AFS (local) и др.
	Организация процессов		Block Report		Black lists
	Мотивационная схема	Анализ портфеля	Анализ fraud-сегментов		Биометрия
	Мониторинг деятельности		Исследования (FTI)	Внешние сервисы	FPS, AFS (межбанк)
Клиенты	Идентификация клиента		Ad hoc		FPS.Bio
	Визуальная оценка	Модели	Предиктивная аналитика		Телеком (МегаФон и др.)
	Андеррайтинг		Разработка правил		Соцсети (Mail.ru и др.)
Партнеры	Партнерский договор		Machine Learning		Cookie (Rambler и др.)
	Проверка партнеров	Оценка	СВА антифрод-процессов		СПАРК, Контур.Фокус и др.
	Мониторинг деятельности		СВА антифрод-систем		Банкроты (Interfax)
Процессы	Аудит процессов		Оценка fraud-потерь	Госсервисы	ФМС
	Согласование инициатив				ФССП
	Автоматизация процессов				ПФР
					ФНС

[1] Стоимость превентивных инструментов отражает не только затраты на их разработку и внедрение, но и потери, связанные с ограничениями на бизнес банка (падение объемов продаж, снижение привлекательности/удобства продукта и т.п.). В принципе это характерно и для директивных инструментов, а в некоторых случаях и для детективных.

Как с помощью аналитики повысить эффективность банковского антифрода?

— бизнес-процессы: персонал, клиенты, партнеры, процессы;

— аналитика: отчетность, портфельный анализ, моделирование, оценка;

— антифрод-системы: внутренние системы, внешние сервисы, государственные сервисы.

Такая карта позволяет держать «на виду» все известные нам антифрод-инструменты и мероприятия, тестировать их, оценивать и принимать решение о необходимости внедрения новых инструментов или о целесообразности отключения уже внедренных.

Стоит отметить, что большую часть рисков мошенничества можно закрыть на уровне «Бизнес-процессы» (причем для этого необязательно нанимать многочисленный штат антифрод-специалистов или сотрудников службы безопасности). Оставшиеся риски можно закрывать с помощью антифрод-систем, внедрение и поддержка которых находятся на стороне вендоров и IT-подразделений банка. Возникает вопрос: для чего тогда в антифроде нужна аналитика? Этому есть несколько очевидных причин (которым на самом деле не всегда уделяется должное внимание):

1. С помощью аналитики можно корректно оценивать эффективность антифрод-решений, предлагаемых вендорами. То, что работает в одном банке, зачастую не работает в другом (по разным причинам: от разной степени защищенности бизнес-процессов и IT-систем до разного клиентского профиля, каналов продаж, банковских продуктов и т.п.).

2. Закрытие рисков мошенничества на уровне бизнес-процессов банка — это тоже расходы (как уже отмечалось, эти расходы отражают не только стоимость разработки и внедрения, но и потери, связанные с влиянием на бизнес). Поэтому для оценки данного типа инструментов аналитика не менее важна (и часто оказывается более сложной, поскольку нет возможности провести ретротест).

3. Аналитическое подразделение отвечает за «in-house разработку» — кастомизированные антифрод-инструменты, внутренняя разработка которых оказывается дешевле аналогичных решений внешних вендоров. К этой тематике можно отнести различные отчеты, мониторинги, кастомизированные модели и правила, исследования и т.д.

Для иллюстрации важности перечисленных причин разберем несколько аналитических примеров из практики работы разных банков.

«Зефирный тест»

В 60-х годах прошлого столетия психологи Стэнфордского университета провели несколько исследований по изучению отложенного удовольствия у детей. Первый эксперимент проводился в 1960 г.

Сергей АФАНАСЬЕВ

в детском саду Стэнфордского университета[1]. Детей в возрасте от четырех до шести лет проводили в комнату без отвлекающих факторов, где на столе стояла тарелка с лакомством, выбранным ребенком заранее (зефир, печенье или крендель). Исследователи разрешали детям съесть лакомство сразу или подождать 15 минут, по истечении которых ребенок получал вторую порцию лакомства в качестве вознаграждения за терпение. Во время эксперимента дети вели себя по-разному. Некоторые прикрывали глаза руками или отворачивались, чтобы не видеть угощение. Другие играли с лакомством как с игрушкой. Некоторые просто съедали зефир или печенье сразу, как только взрослые уходили из комнаты. В эксперименте приняли участие более 600 детей. Меньшинство съели лакомство сразу, примерно треть детей вытерпели 15 минут и получили двойное угощение.

Когда дети выросли, ученые провели новые исследования и обнаружили, что у более терпеливых детей из «зефирного теста» жизнь складывалась лучше по разным показателям: по результатам школьных тестов, по уровню образования, по индексу массы тела и т.д.[2]

Отложенные мошеннические схемы маловероятны. Поэтому когда проводится поиск уязвимостей в бизнес-процессах банка и рассматриваются гипотетически возможные мошеннические схемы, в первую очередь необходимо прорабатывать «быстрые» схемы.

Если использовать эти выводы в антифроде, то можно предположить, что мошенники менее терпеливы, чем добропорядочные люди. И скорее всего, мошенники в детстве съедали все самое вкусное сразу, а потом хуже учились, хуже справлялись с рабочими обязанностями, хуже росли по карьерной лестнице и т.д. Но при этом у мошенников оставались амбиции получать все и сразу.

Доказательством этой гипотезы могут служить полученные нами результаты оценки эффективности ревизии POS-партнеров банка. Цель данной ревизии — выявить POS-партнеров банка, которые долго не выдавали кредиты. Обычно таких партнеров блокируют, чтобы минимизировать риски отложенного мошенничества. Чтобы оценить эффективность этой процедуры, мы разбили портфель POS-кредитов на две группы:

1) кредиты, выданные POS-партнерами до длительного перерыва в работе;

2) кредиты, выданные POS-партнерами после длительного перерыва в работе.

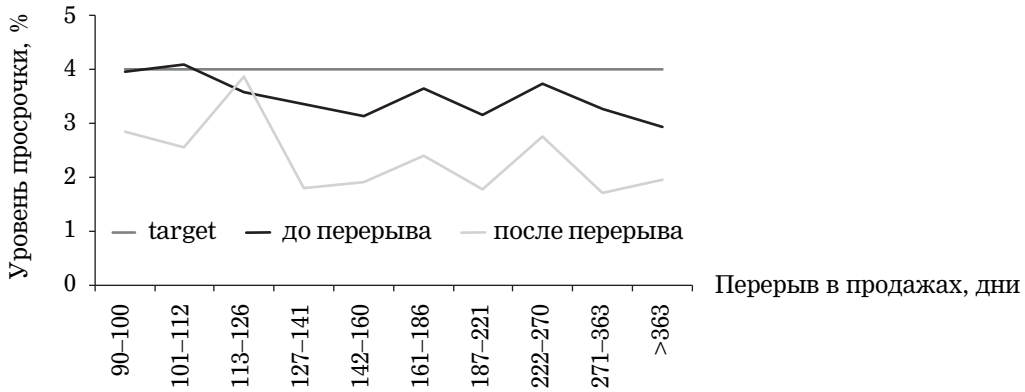
В качестве «длительного перерыва» мы брали разные периоды: от 3 месяцев до более 1 года. Полученные результаты показаны на рис. 3.

[1] Mischel W., Ebbesen E.B. Attention in delay of gratification // Journal of Personality and Social Psychology. 1970. Vol. 16. No. 2. P. 329–337.

[2] Shoda Y., Mischel W., Peake P.K. Predicting Adolescent Cognitive and Self-Regulatory Competencies from Preschool Delay of Gratification: Identifying Diagnostic Conditions // Developmental Psychology. 1990. Vol. 26. No. 6. P. 978–986.

Рисунок 3

Зависимость просрочки от длительности перерыва в продажах POS-партнеров



Оказалось, что кредиты, выданные после длительного перерыва, более качественные. Просрочка по таким кредитам в среднем в 1,5 раза ниже, чем по кредитам, выданным до длительного перерыва. Это интересные результаты, которые, во-первых, показали, что наши предположения о рисках отложенного мошенничества были неверными (т.е. нельзя блокировать всех партнеров, которые долго не продавали)[1]. А во-вторых, эти результаты наглядно демонстрируют, как работает «зефирный тест»: мошенники не будут долго ждать, если у партнера есть предрасположенность к мошенничеству — обычно он ворует сразу.

Из «зефирного теста» можно сделать общий вывод, что отложенные мошеннические схемы маловероятны. Поэтому когда проводится поиск уязвимостей в бизнес-процессах банка и рассматриваются гипотетически возможные мошеннические схемы, в первую очередь необходимо прорабатывать «быстрые» схемы, которые будут чаще встречаться и использоваться мошенниками.

Бином Ньютона

Большинство современных аналитических инструментов не появились бы на свет без теории вероятностей и математической статистики. Например, с помощью бинома Ньютона можно оценить веро-

[1] Это наглядный пример принципов, которые продвигал известный американский аудитор Роберт Монтгомери (считающийся в профессиональных кругах «дедушкой» современного аудита). Монтгомери высказывал мысль, что аудит должен проводиться в интересах организации, ее собственников, клиентов и других инвесторов, а не ради искоренения ошибок и нарушений, как это делали ранее. Труды Монтгомери изложены в книге «Аудит Монтгомери» (старое название: «Аудит: теория и практика»).

Сергей АФАНАСЬЕВ

ятность наступления того или иного события, пронаблюдав последовательность нескольких независимых случайных событий. Формула Байеса позволяет определить вероятность какого-либо события при условии, что уже произошло другое взаимозависимое с ним событие. Метод максимального правдоподобия позволяет проводить оценки на фиксированных выборках и обобщать полученные результаты на всей совокупности объектов. В антифрод-задачах мы часто сталкиваемся с этими понятиями, когда нам необходимо оценить вероятность реализации мошеннических схем или обобщить результаты построенных фрод-моделей.

К примеру, у нас в банке реализован процесс автоматической блокировки сотрудников и партнеров по высоким рисковым показателям: если показатели просрочки по сотруднику (или партнеру) превышают пороговые значения, то логин этого сотрудника (или партнера) автоматически блокируется.

Когда мы настраивали критерии блокировок по этому процессу, перед нами встал вопрос: какое количество дефолтных кредитов по сотруднику (или партнеру) считать статистически значимым? Например, если сотрудник выдал один кредит и этот кредит оказался в дефолте, то уровень просрочки по этому сотруднику будет равен 100% — считать ли это случайностью? Наверное, да. А если сотрудник выдал два кредита и оба оказались в дефолте — это случайность? Возможно, совпадение. А если три кредита из трех оказались в дефолте — это все еще случайность или уже закономерность? Ответ на этот вопрос дает бином Ньютона.

Общая формулировка задачи звучит следующим образом: *какова вероятность того, что из n кредитов k и более кредитов окажутся в просрочке случайно, при среднем уровне просрочки по портфелю d?*

Вероятность такого события считается по формуле бинома Ньютона:

$$P = \sum_{i=k}^{i=n} C_n^i d^i (1-d)^{n-i}. \quad (1)$$

В наших расчетах средний уровень просрочки по портфелю был равен 3,7%. Можно, например, оценить, какова вероятность того, что два из двух кредитов случайно окажутся в просрочке (здесь $k = 2$, $n = 2$, $d = 0{,}037$). Подставив все исходные значения параметров в формулу (1), получаем, что $P = 0{,}0014$. То есть если сотрудник выдал два кредита и оба этих кредита оказались в дефолте, вероятность такого события составляет всего 0,14% (если выпадение в просрочку этих кредитов считать случайностью). На рис. 4а показаны аналогичные расчеты для сотрудников, выдавших не более пяти кредитов

и допустивших хотя бы одну просрочку (в верхней строке указано количество выданных кредитов, в левом столбце — минимальное количество дефолтных кредитов, а в ячейках — вероятности P).

Чтобы проверить, как теория работает на практике, мы посмотрели, как сотрудники, показанные на рис. 4а, вели себя после реализации этого события. То есть мы посчитали просрочку по кредитам, которые были выданы сотрудниками после расчетного периода, используемого на рис. 4а. Полученные результаты представлены на рис. 4б, где в верхней строке и левом столбце те же обозначения, что и на рис. 4а, а в ячейках указан уровень просрочки по сотрудникам за 12 месяцев после расчетного периода.

На рис. 4а заливкой обозначены диапазоны вероятностей P:

— белая заливка: $P > 5\%$;

— серая заливка: $1\% < P \leq 5\%$;

— черная заливка: $P \leq 1\%$.

На рис. 4б заливкой обозначены диапазоны просрочки относительно целевого значения просрочки $T = 4{,}8\%$ (T — нулевой таргет, или точка безубыточности):

— белая заливка: просрочка ниже T;

— серая заливка: просрочка выше T, но ниже $2T$;

— черная заливка: просрочка выше $2T$.

Видно, что цвета между рис. 4а и 4б сильно коррелируют, то есть сотрудники, попавшие в диапазон вероятностей $P \leq 1\%$ (событие маловероятно), оказались потом убыточными для банка и в следу-

Рисунок 4

Вероятность и реальный уровень просрочки: (а) — вероятность случайного попадания кредитов в просрочку (при средней просрочке по портфелю d = 0,037); (б) — просрочка по сотрудникам за 12 месяцев после расчетного периода (%)

(а)

k \ n	1	2	3	4	5
1+	3,7	7,3	10,7	14,0	17,2
2+		0,14	0,40	0,78	1,27
3+			0,01	0,02	0,05
4+				0,00	0,00
5					0,00

(б)

k \ n	1	2	3	4	5
1+	6,1	5,6	5,6	4,9	5,2
2+		7,1	7,1	5,4	6,9
3+			13,4	7,7	10,5
4+				10,0	22,2
5					56,3

Сергей АФАНАСЬЕВ

ющие 12 месяцев работы выдавали кредиты с просрочкой, значительно превышающей целевые значения банка.

После проверки формул на реальной статистике можно смело использовать рассчитанные по биному Ньютона вероятности P в качестве корректирующих коэффициентов для определения порогов автоматической блокировки. В результате из расчетов не будут исключены сотрудники и партнеры с «маленькими продажами», которые генерируют банку убыток.

Парадокс дней рождений

Прежде чем приступить к рассмотрению последнего кейса, разберем еще одну задачу из теории вероятностей, которую называют «парадокс дней рождений». Эта задача обычно дается на семинарах по курсу теории вероятностей и звучит так: *какова вероятность того, что в группе из 23 человек хотя бы у двух совпадут дни рождения (число и месяц)*[1]?

Всегда надо аккуратно относиться к результатам тестирований сервисов фотобиометрии, голосовой биометрии и т.д. Тестирование обычно проводится на маленьких выборках, а в промышленной эксплуатации на больших данных может реализоваться «проклятье дней рождений».

Предполагается, что в группе студентов, которым дают эту задачу, все одного года рождения и численность группы составляет примерно 23 человека. В общей постановке задачи также предполагается, что дни рождения распределены равномерно (в году 365 дней, нет близнецов, рождаемость не зависит от дня недели, времени года и других факторов).

В этой задаче довольно часто интуитивное восприятие строится на том, что вероятность совпадения дней рождения у двух человек с любым днем в году составляет $1/365 = 0{,}27\%$. Эта вероятность умножается на количество человек в группе (23) и получается итоговая вероятность: $(1 / 365) \times 23 = 6{,}3\%$. Однако такие рассуждения неверны, так как число возможных пар составляет: $(23 \times 22) / 2 = 253$, что значительно превышает количество человек в группе.

Чтобы вывести общую формулу, необходимо рассчитать вероятность того, что в группе из n человек ($n \leq 365$) дни рождения у всех людей будут различными (обозначим эту вероятность через q). Возьмем случайным образом любого человека из группы и запомним его день рождения. Затем возьмем наугад второго человека. Тогда вероятность того, что день рождения первого не совпадет с днем рождения второго, будет равна: $1 - (1/365)$. Если взять наугад третьего человека, вероятность того, что его день рождения не совпадет с днями рождения первых двух, будет равна: $1 - (2/365)$. И так далее до последнего, для которого вероятность несовпадения дня рожде-

[1] Козлов М.В. Элементы теории вероятностей в примерах и задачах. МГУ, 1990.

ния с днями рождения всех остальных будет равна: $1 - ((n - 1) / 365)$. Перемножая эти вероятности, получим вероятность того, что в группе из n человек ($n \le 365$) у всех людей будут разные дни рождения:

$$q = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right)\ldots\left(1 - \frac{n-1}{365}\right) = \frac{365!}{365^n (365 - n)!}. \quad (2)$$

Тогда по формуле обратной вероятности получим вероятность того, что хотя бы у двух человек дни рождения совпадут:

$$p = 1 - \frac{365!}{365^n (365 - n)!}. \quad (3)$$

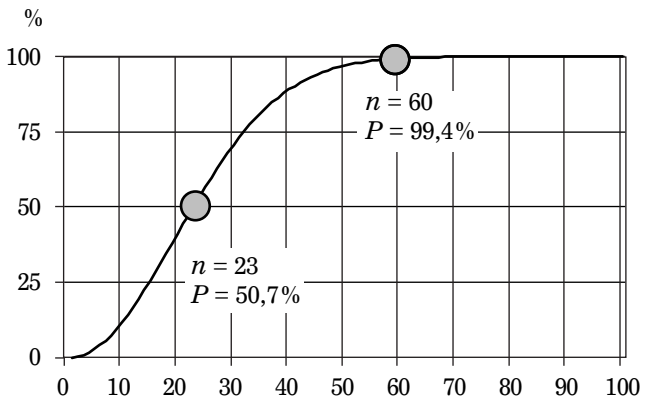
Для $n = 23$ вероятность $p = 50{,}7\%$, а для группы из 60 человек вероятность вообще становится более 99%, то есть значительно отличается от нашего интуитивного восприятия (рис. 5).

Парадокс дней рождений часто используется в практических задачах, например, в криптоанализе этот эффект применяется в методах для взлома шифров и поиска коллизий хэш-функций (т.н. «атака дней рождений»)[1].

Мы с этим эффектом столкнулись несколько лет назад, когда внедряли в банке локальную биометрическую систему, обрабатывающую фотографии. Основное назначение данной системы было в сравнении фотографий из кредитных заявок с общей базой клиентских фотографий (поиск похожих лиц). Соотвественно если био-

Рисунок 5

Вероятность совпадения даты рождения хотя бы у двух человек в группе



[1] Bellare M., Kohno T. Hash Function Balance and Its Impact on Birthday Attacks. EUROCRYPT 2004.

Сергей АФАНАСЬЕВ

метрическая система находила фотографии с одинаковыми лицами и при этом персональные данные этих клиентов отличались (ФИО + дата рождения + паспорт), то можно было сказать, что с высокой вероятностью один из этих клиентов — мошенник, использующий поддельный паспорт.

Общий процесс работы биометрической системы выглядел следующим образом:

1. Заявитель обращался в банковский офис за кредитом.

2. Сотрудник банка заполнял заявку и фотографировал клиента. Данные отправлялись в скоринговую и биометрическую системы.

3. Биометрическая система сравнивала фотографию заявителя с каждой фотографией из клиентской базы и рассчитывала коэффициенты похожести (скорбаллы).

4. Топ-30 фотографий с наивысшими коэффициентами похожести отправлялись на проверку андеррайтеру.

5. Андеррайтер сравнивал фотографию заявителя с 30 похожими фотографиями из базы и отмечал те, на которых, по его мнению, был один и тот же человек.

6. Если по заявителю находились одинаковые фотографии из топового пула с другими персональными данными (ФИО + дата рождения + паспорт), то заявка отклонялась как мошенническая.

Тестирование биометрической системы по описанной схеме показало хорошие результаты. Было собрано 10 тыс. фотографий, точность биометрической системы на этой базе оказалась достаточно высокой, чтобы запустить систему в промышленную эксплуатацию[1].

После ввода биометрии в промышленную эксплуатацию размер базы стал расти. Когда количество фотографий достигло 1 млн, мы пересчитали значения показателей точности и очень удивились, когда увидели, что эти значения упали практически до нуля. Убедившись, что расчеты верны, нам надо было разобраться в причинах такого сильного падения точности. А поняв эти причины, придумать решение — как повысить точность до уровня тестовых значений.

На помощь пришла теория вероятностей, а именно задача о «парадоксе дней рождений». Рассуждения были следующими: если в формуле для дней рождений заменить количество человек в группе (n) на количество фотографий в топовом пуле (30), а количество дней в году — на количество фотографий в общей базе, то принципы

[1] Стоит отметить, что это происходило в те времена, когда еще не было современных архитектур для сверточных нейронных сетей и биометрические системы работали на примитивных алгоритмах машинного обучения, показывая в среднем не очень высокую точность.

Как с помощью аналитики повысить эффективность банковского антифрода?

работы формулы переносятся на нашу задачу. То есть если количество фотографий в общей базе будет небольшим (близким к количеству дней в году), вероятность случайного попадания одинаковых лиц в топовый пул (случайно выбранных из базы 30 фотографий) будет достаточно высокой.

Таким образом, можно оценить, как меняется теоретическая точность[1] биометрической модели в зависимости от размера базы фотографий и ряда других параметров. Общая формула для расчета теоретической точности получается следующей:

$$P = 1 - \left(1 - \frac{C_{N-1}^{T} - C_{N-K}^{T}}{C_{N-1}^{T} N} MK\right)^{R}, \quad (4)$$

где N — общее количество фотографий в базе;

T — количество фотографий в топовом пуле;

K — количество фотографий в базе на одного мошенника с разными паспортными данными (экспериментально $K = 6$);

R — количество регионов, на которые сегментирована база (если база не сегментирована, то $R = 1$);

M — количество мошенников в сегменте.

Посчитав значения теоретической точности для различных параметров, мы увидели, что с увеличением размера базы теоретическая точность сильно падает (рис. 6). При этом справедливо и обратное утверждение: с уменьшением базы теоретическая точность растет.

Таким образом, нам стало понятно: чтобы повысить точность, нужно сегментировать базу фотографий на несколько маленьких кусочков и запускать биометрию отдельно на каждом кусочке. Из практики мы видели, что мошенники, оформляющие кредиты по переклеенным паспортам, как правило, работали в пределах одного города. То есть всю базу фотографий можно было сегментировать по регионам (1 000 000 / 84 ≃ 12 000). Кроме того, можно было сделать гендерную сегментацию и разделить каждый кусочек еще пополам (12 000 / 2 = 6000). Такой сегментации оказалось более чем достаточно, чтобы биометрия заработала с высокой точностью.

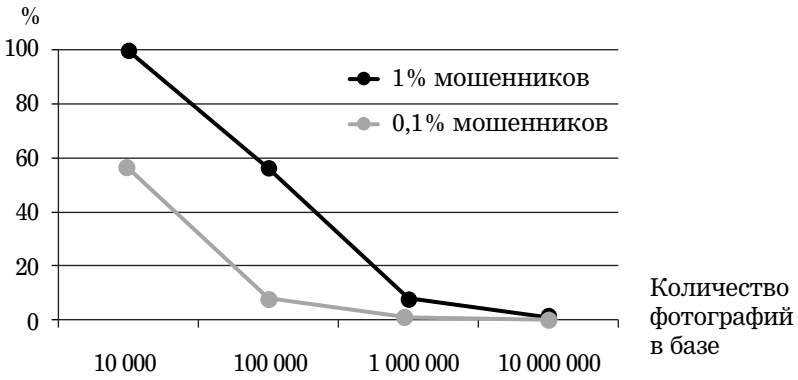
Основной урок, который мы получили из этого кейса, — всегда надо аккуратно относиться к результатам тестирований подобных сервисов (фотобиометрия, голосовая биометрия, отпечатки пальцев и т.д.). Тестирование обычно проводится на маленьких выборках,

[1] Под теоретической точностью понимается вероятность случайного попадания похожей фотографии в топовый пул, то есть делается предположение, что биометрическая модель работает в режиме «монетки», формируя топовый пул из случайно выбранных фотографий.

Сергей АФАНАСЬЕВ

Рисунок 6

Вероятность попадания похожей фотографии в пул из 30 случайно выбранных фотографий из общей базы



а в промышленной эксплуатации на больших данных может реализоваться «проклятье дней рождений»[1].

Итак, мы рассмотрели несколько примеров, демонстрирующих, как с помощью аналитики можно повысить эффективность антифрод-инструментов, мероприятий и принимаемых решений. И здесь, наверное, необходимо вспомнить цитату, которую приписывают американскому бейсболисту Йоги Берре[2]: «В теории нет разницы между теорией и практикой. А на практике есть».

Очевидно, что и в антифроде есть большая разница между теорией и практикой. Отчасти это связано с тем, что инструментарий для антифрода мы берем из разных дисциплин (аудита, криминалистики, информационной безопасности и т.д.), и часто нам просто не хватает знаний в этих областях. Другой причиной является то, что мошенники быстро адаптируются, придумывая новые схемы, в ответ на которые появляются новые антифрод-инструменты. В результате управлять таким огромным инструментарием становится крайне сложно. Тем не менее, разница между теорией и практикой — это те фрод-потери банка, которые можно предотвратить, если сократить данный разрыв. И мы надеемся, что с помощью аналитики этого можно добиваться.

[1] Стоит отметить, что во многих современных биометрических решениях используются подобные техники сегментаций.

[2] Лоуренс Питер (Йоги) Берра (1925–2015) — один из величайших американских бейсболистов, который был известен своими «йогизмами» — комментариями и остротами, имеющими форму очевидной тавтологии или парадоксального противоречия. Некоторые «йогизмы» часто приписывались Берре, хотя он их никогда публично не произносил.