

Методический сборник

**Операционный риск:**

как перестроить систему управления и выполнить требования 716-П

**Операционный риск:  
как перестроить систему  
управления и выполнить  
требования 716-П**

Объем: 112 с.

Формат: А4

Издано в декабре 2021 г.

Банковскую отрасль накрыло волной новых регуляторных требований к управлению операционным риском. В 2020 г. вышли Положения Банка России № 716-П и № 744-П, в 2021-м точка не поставлена, и банки ждут изменений и уточнений.

В сборник вошли статьи, в которых анализируются яркие события операционного риска с точки зрения риск-менеджмента, представлен первый опыт банков по трансформации СУОР, даны ответы экспертов на вопросы аудиторов и риск-менеджеров.

**В числе тем:**

- что придется перестроить из-за регуляторной трансформации?
- что будет проверять регулятор и как новые требования скажутся на капитале банков?
- какие факторы в наибольшей степени влияют на реализацию операционного риска?
- что нужно для выполнения требований Положения № 716-П, кроме автоматизации и сбора базы данных?
- какие инструменты автоматизации использовать для мониторинга операционного риска и его оценки?
- как сделать бизнес-подразделения своими союзниками и мотивировать сообщать об инцидентах?
- какие внешние данные учитывать при моделировании?
- по какой методике аудиторам проводить оценку эффективности управления операционным риском?

**Скачать ознакомительный фрагмент:**[http://www.reglament.net/pages/acquaintance\\_demo.htm?id=154](http://www.reglament.net/pages/acquaintance_demo.htm?id=154)

Реклама

**Риск-менеджмент  
в кредитной организации**

Методический журнал

**№ 4 (44) \ 2021**'21  
4

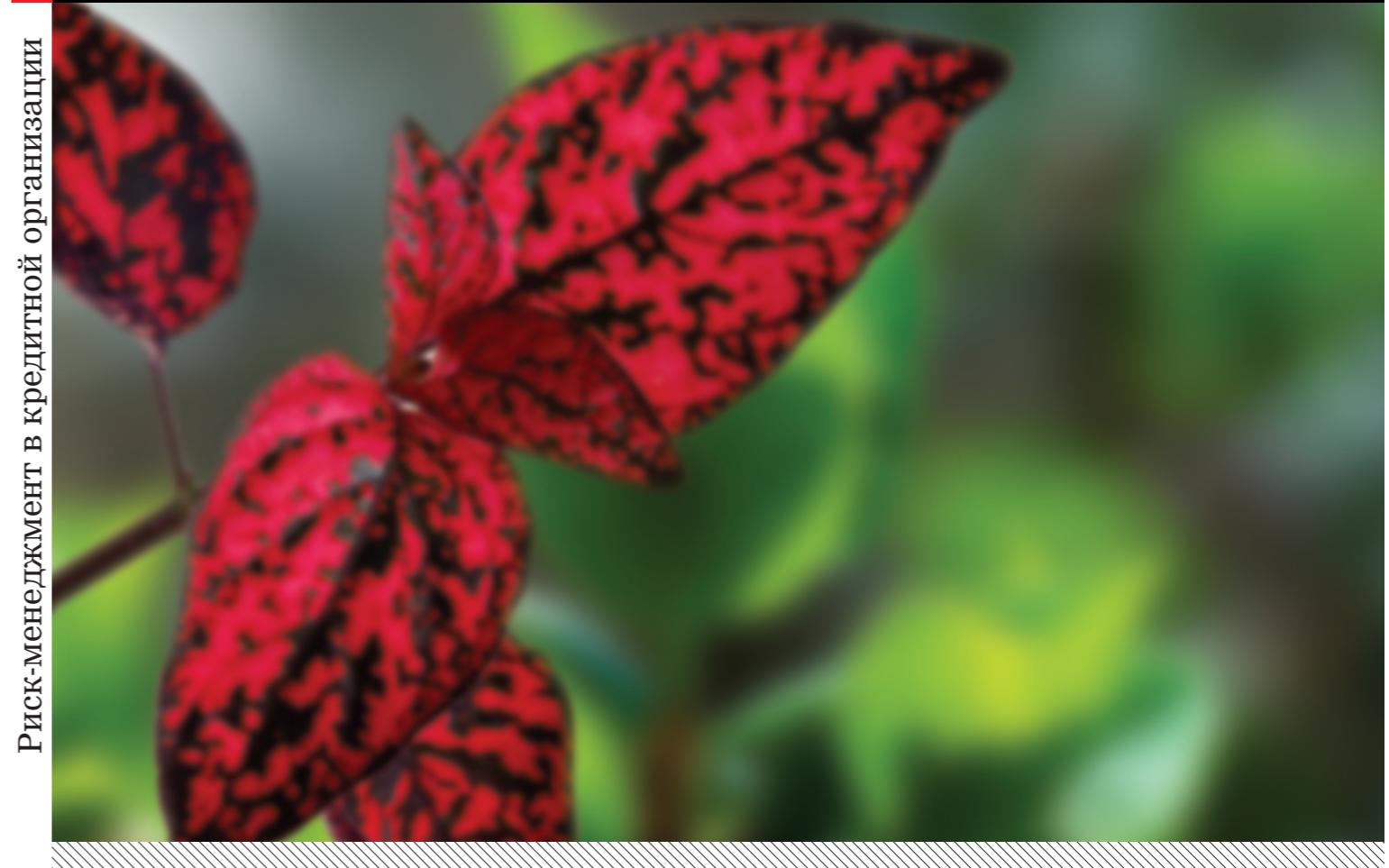
РЕГЛАМЕНТ®

**Модели в банках: возможности, угрозы, риски**

Лучшие практики data science проектов для управления модельным риском: опыт QIWI

Алексей Лобанов: итоги 10 лет внедрения Базеля

LGD-модели для розничного кредитования. Часть 2: разработка «ядра» модели



Методический сборник

## Антикризисные методики

Часть 1

Антикризисные методики. Как оценить реальное положение дел и снизить негативное влияние ситуации на бизнес банка

Объем: часть 1 — 134 с., часть 2 — 159 с.

Формат: А4

Издано: часть 1 – июнь 2020 г., часть 2 – ноябрь 2020 г.

По мнению экспертов, в ближайшей перспективе отчетность банков не будет отражать реального положения дел. Послабления, введенные Банком России, позволяют не только скрыть реальные потери, но даже повысить рентабельность операций на фоне экономического кризиса, вызванного всеобщей пандемией. В этот период профильные подразделения в банках должны разрабатывать и внедрять неотложные меры реагирования. Чтобы облегчить эту задачу, мы публикуем антикризисные методические разработки от ведущих банковских специалистов.

Лучшие материалы 2020 г. объединены в тематические разделы: «Риски кредитования», «Бизнес-процессы», «Форс-мажор», «Финансовые инструменты», «Кредитный скоринг», «Аудит и комплаенс», «Возврат проблемной задолженности».

В ноябре 2020 г. вышла вторая часть сборника, из которой вы узнаете, как адекватно оценивать не-предвиденные кредитные потери в условиях пандемии, выдавать ли потребительские кредиты тем, кто мог пострадать от кризиса, как сделать шаг навстречу хорошим корпоративным заемщикам и какими должны быть три основных вектора аудиторских проверок в условиях волатильности уровней кредитного, рыночного и операционного рисков.

По ссылке [http://www.reglament.net/bank/am\\_2020/](http://www.reglament.net/bank/am_2020/) вы можете скачать ознакомительный фрагмент, в который вошли 3 материала:

- ▶ **Владимир КОЗЛОВ**, FRM, консультант по риск-менеджменту, [raisk.ru](#)  
Свежие идеи для розничного кредитования юридических лиц: где их взять и как их оцифровать?
- ▶ **Александр КАРПОВ**, банковский эксперт  
Инструменты бизнес-аналитики для кредитного анализа корпоративного клиента
- ▶ **Антон МЕЛЕНЦОВ, Александр ДЕНИСОВ**, ООО «Сервис-модель» (г. Екатеринбург)  
Комплексный анализ и моделирование ключевых показателей сети кэш-поинтов в различных сценарных условиях

Реклама

Подробнее на сайте [www.reglament.net](http://www.reglament.net)  
и по телефону: +7 (495) 255-5177

Методический сборник

## Кредитный риск: как найти «точку отказа»

Методика

«точку отказа»

Методический сборник

## Кредитный риск: как найти «точку отказа»

Подходы банковских практиков к оценке заемщика

Объем: 149 с.  
Формат: А4  
Издан в июне 2021 года

Уровень одобрения/отказа в кредитовании должен быть экономически оправдан для банка. По какой методике искать диапазон оценок, подтверждающих экономическую оправданность отказа в кредитной сделке? Как настроить challenger-стратегию для оценки эффективности кредитного конвейера? Как управлять специфическими рисками контрактного кредитования и каковы особенности оценки заемщиков из разных отраслей?

Ответы на эти и многие другие вопросы — в новом сборнике «Кредитный риск: как найти «точку отказа»». Над материалами в нем потрудились ведущие практики из крупных и средних банков, ежедневно меняющие топографию банковского бизнеса, а также эксперты, разрабатывающие методики банковского кредитования.

### Рубрики:

- ▶ Риски розничного кредитования. Скоринг
- ▶ Оценка корпоративного заемщика
- ▶ Отраслевая специфика
- ▶ Инструменты анализа

### Авторы материалов:

- ▶ Михаил ПОМАЗНОВ, ПАО «Промсвязьбанк», Дирекция рисков, руководитель по валидации
- ▶ Елена КОНЕВА, FICO, директор региона EMEA
- ▶ Павел НИКОЛАЕВ, Банк «Открытие», управляющий директор департамента интегрированных рисков
- ▶ Александр БОРОДИН, GlowByte Consulting, ведущий эксперт по моделированию финансовых рисков
- ▶ Алексей БУЗДАЛИН, Группа «Интерфакс», директор Центра экономического анализа
- ▶ Денис СКРИПКА, компания Digital Agro, руководитель разработки продуктов
- ▶ и другие

Реклама

Подробнее на сайте [www.reglament.net](http://www.reglament.net)  
и по телефону: +7 (495) 255-5177

# Риск-менеджмент в кредитной организации

## Методический журнал

Издаётся с 2011 года.  
Выходит один раз в квартал

№ 4 (44) \ 2021

Зарегистрирован Федеральной службой по надзору  
в сфере связи, информационных технологий  
и массовых коммуникаций (Роскомнадзор)  
1 июля 2010 года.  
Свидетельство о регистрации ПИ № ФС77-40479

Учредитель и издатель  
**ООО «Регламент»**  
www.reglament.net

Генеральный директор **В.Г. Богданов**

Главный редактор **В.С. Козлов**  
kozlov@reglament.net

Ответственный секретарь Департамента  
финансовых и методических изданий  
**И.М. Ананьева**  
ananieva@reglament.net  
Выпускающий редактор **Е.В. Полякова**

**Отдел предпечатной подготовки  
и производства**  
Начальник отдела **А.Н. Тимченко**  
Верстка **С.В. Шеришорин**

**Отдел маркетинга**  
Директор по маркетингу **А.В. Гришуин**  
grishunin@reglament.net

© ООО «Регламент», 2021

**Индексы в каталогах**  
Роспечать: 36193  
УП УРАЛ-ПРЕСС: 36193  
«Книга-Сервис»: 26167

**Подписка через Интернет**  
www.reglament.net

**Редакционная подписка**  
возможна с любого месяца.  
Телефон отдела прямых продаж  
(495) 255-5177, доб. 215  
E-mail: podpiska@reglament.net

По всем вопросам, связанным с доставкой изданий и отчетных документов, обращайтесь в отдел распространения и логистики ООО «Регламент» по тел. (495) 255-5177, доб. 289.

Мнения, оценки и рекомендации в статьях, размещенных в журнале, отражают точку зрения их авторов и не являются обязательными к исполнению. ООО «Регламент» и авторы материалов, опубликованных в журнале, не несут ответственности за возможные убытки, которые могут быть причинены лицам в результате использования или невозможности использования ими размещенных материалов. Пользователь самостоятельно оценивает возможные риски совершения юридически значимых действий на основе размещенной в журнале информации и несет ответственность за их неблагоприятные последствия. Полное или частичное воспроизведение каким-либо способом материалов, опубликованных в журнале, допускается только с письменного разрешения редакции. Редакция не несет ответственности за достоверность информации в рекламных объявлениях.

Адрес учредителя, издателя и редакции: 125167, Ленинградский просп., 37, БЦ «Аэродом», 8 этаж, оф. 8.2  
Телефон (495) 255-5177.

Отпечатано в ООО «КЛУБ ПЕЧАТИ». Адрес: 127018, Москва, 3-й проезд Марьиной Рощи, 40, стр. 1, оф. 32.  
Тираж 1500 экз. Цена свободная. Подписано в печать 24.12.2021.

### Экспертный совет журнала

**Сергей АФАНАСЬЕВ**, КБ «Ренессанс Кредит» (ООО), вице-президент, начальник управления статистического анализа

**Александр ДЬЯКОНОВ**, профессор ВМК МГУ, д. ф.-м.н.

**Сергей КАПУСТИН**, Азиатско-Тихоокеанский Банк, заместитель председателя правления

**Алексей ЛОБАНОВ**, Банк России, Департамент банковского регулирования, экс-директор

**Игорь ФАРРАХОВ**, ООО «РИСКФИН», заместитель генерального директора

## Содержание

### ФИНАНСОВЫЕ КАТАСТРОФЫ. ОПЕРАЦИОННЫЙ РИСК

6 Редакция журнала

#### **ПЕРВАЯ КРУПНАЯ АТАКА С ИСПОЛЬЗОВАНИЕМ DEEP FAKE В БАЗЕ ORX: ЧТО ДАЛЬШЕ?**

В базу данных ORX впервые попал крупный инцидент мошенничества с использованием AI. Технология deep voice позволила злоумышленникам вывести из компании \$35 млн. Какие методы помогут банкам противостоять таким атакам и что должно быть приоритетной задачей при защите от них?

### КРЕДИТНЫЙ РИСК

16 **АЛЕКСЕЙ ЛОБАНОВ: ИТОГИ 10 ЛЕТ ВНЕДРЕНИЯ БАЗЕЛЯ**

Спустя 10 лет с того момента, как Алексей Лобанов начал карьеру в Банке России, и через 5 лет после назначения директором Департамента банковского регулирования мы задали ему вопросы наших коллег-рисковиков: что удалось сделать за эти годы и почему требования Банка России столь тяжелы для восприятия и исполнения?

22 Сергей АФАНАСЬЕВ, Анастасия СМИРНОВА, Ильгиз АХМЕТСАФИН,  
Игорь МОЛОКАНОВ, КБ «Ренессанс Кредит»

#### **LGD-МОДЕЛИ ДЛЯ РОЗНИЧНОГО КРЕДИТОВАНИЯ.**

#### **ЧАСТЬ 2: РАЗРАБОТКА «ЯДРА» МОДЕЛИ**

В предыдущей статье мы разобрали первые три фазы проекта разработки модели LGD согласно стандарту CRISP-DM. Разработка модели является следующей фазой, которую в рамках методологии ПВР и МСФО (IFRS) 9 можно разделить на три этапа: разработка «ядра» модели, калибровка модели, применение надбавок. Рассмотрим первый этап.

39 Сергей КОПЫЛОВ, ООО «Бизнес Системы Консалт»

#### **ЕЖЕКВАРТАЛЬНАЯ ОТЧЕТНОСТЬ РОЗНИЧНЫХ КРЕДИТНЫХ ПОРТФЕЛЕЙ: КАК СОЗДАТЬ ЦЕННОСТЬ БИЗНЕСУ? ЧАСТЬ 2**

В прошлом номере мы рассматривали ежемесячные отчеты, в этом сосредоточимся на отчетах ежеквартальных и немного на макроэкономике. Приведенные в статье примеры отражают состояние реальных кредитных портфелей и позволяют оценить влияние эпидемии COVID-19.

51 Редакция журнала

#### **ОТ СПАРК ДО CREDITNET: ГЛУБИННЫЙ ОБЗОР СИСТЕМ АНАЛИЗА КОНТРАГЕНТОВ. «ЗАЧЕСТНЫЙБИЗНЕС»**

В банковской системе есть сервисы ниже «ватерлинии», которыми почему-то не принято пользоваться. Подключим сервис интеграции «ЗаЧестныйБизнес» по цене ежемесячного утреннего кофе и посмотрим: может, все-таки дать шанс бюджетным решениям и сэкономить миллионы рублей в месяц?

# Риск-менеджмент в кредитной организации

## Методический журнал

№ 4 (44) \ 2021

### АНАЛИЗ ДАННЫХ

59 Лидия ХРАМОВА, QIWI

#### **ЛУЧШИЕ ПРАКТИКИ DATA SCIENCE ПРОЕКТОВ ДЛЯ УПРАВЛЕНИЯ МОДЕЛЬНЫМ РИСКОМ: ОПЫТ QIWI**

Компания QIWI активно развивает бизнес, основанный на аналитике больших данных. Расскажем о топ-6 факторов, влияющих на модельный риск, и трех китах прозрачного процесса разработки. А также о том, как при помощи минимального набора методов построить модели с life time более двух лет.

64 Редакция журнала

#### **ПОЛНОЦЕННЫЙ ML В МОДЕЛЯХ IRB: ПРОЩАЙ, ЛОГРЕГ?**

Европейская служба банковского надзора начинает сближение ML-моделей с требованиями к расчету риска на капитал. Так как Россия часто впереди планеты всей в имплементации продвинутых практик регулирования, дата-саентистам банков, возможно, стоит озабочиться подготовкой к изменениям.

### РЫНОЧНЫЙ РИСК

74 Эмилио ЛЬОРЕНТЕ, Recognition Asset Management Solutions

#### **КАК С ПОМОЩЬЮ MATLAB РАЗМЕСТИТЬ ДЕНЬГИ В АКТИВЫ И СНИЗИТЬ РЫНОЧНЫЕ И ОПЕРАЦИОННЫЕ РИСКИ**

Разработанная на основе MATLAB система призвана заменить решения инвестиционного комитета data-driven решениями. Команда RAMS изучила мнение каждого эксперта, заменила его моделью и снабдила рыночными данными на основании Мертон-моделей. Заметят ли члены комитета разницу?

81 Артем ДАНИЛИШИН, Промсвязьбанк

#### **ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ ОЦЕНКИ VAR/CVAR ПОРТФЕЛЯ ОПЦИОННЫХ КОНТРАКТОВ: НОВЕЙШИЕ ПОДХОДЫ**

Как совместить риск-нейтральную меру ARIMA-GARCH с необходимостью учета тяжелых хвостов? В статье показана действующая модификация расширенного принципа Гирсанова, в котором вместо логарифмических приращений берутся относительные, и приведены примеры кода в среде R.

### МИРОВЫЕ ПРАКТИКИ

85 Редакция журнала

#### **ИССЛЕДОВАНИЕ КРЕДИТНЫХ РЕЙТИНГОВ КОМПАНИЙ КНР: КАЧЕСТВО И ИНТЕРПРЕТАБЕЛЬНОСТЬ**

В КНР экономисты и датасаентисты, похоже, наконец пришли к согласию. Они выработали методологию и тип моделей, которые, во-первых, дружат между собой, а во-вторых, не встречаются ни на Github, ни на Kaggle. Посмотрим, в чем состоял смысл многолетней исследовательской работы в области кредитных рисков и анализа данных.

---

## Содержание

---

### CASE STUDY

#### 90 КАК ОЦЕНИТЬ РИСКИ КРЕДИТОВАНИЯ НЕБОЛЬШОГО ИТ-БИЗНЕСА

Обсудим с менеджером проектов и портфелей Данилом Динцисом, как кредитовать ИП и небольшие компании, которые оказывают услуги в сфере ИТ. Если клиент — ИП, банки часто ограничиваются оценкой финансового положения и предлагают потребкредит. Но для таких клиентов можно создать отдельный продукт, сбалансированный по риску.

### ПЕРСОНАЛЬНАЯ ЭФФЕКТИВНОСТЬ

#### 103 Владимир КОЗЛОВ, FRM

#### SQL-JOIN: РАЗБИРАЕМСЯ ОДИН РАЗ И НАДОЛГО

По итогам нашего анализа hh.ru, в требованиях работодателя SQL занимает место выше Python. А учитывая, что в рисках становится все больше данных и методы работы с ними должен знать каждый — мы, не претендуя на изложение продвинутых подходов, расскажем о том, что пока известно не всем и в чем некоторые допускают ошибки.

# Риск-менеджмент в кредитной организации

## Методический журнал

---

### Слово редактору

---



Владимир КОЗЛОВ, главный редактор, *FRM*

Однажды молодого риск-аналитика пригласили на встречу с клиентом, заключившим госконтракт на несколько миллиардов. Тон беседы задавали оптимизм клиентского подразделения и энтузиазм заемщика. Звучали слова «гарантии», «раскрытие», «поручительства», «залоги», «миллионы», «обслуживание». Рисковик между тем думал, что так и не научился Питону и что ОПГ — это не преступная группировка, как показалось члену совета, а метод обобщенного градиента в Excel. Также его смущали стычки с клиентщиками и необходимость сглаживать углы — ведь его работа в том числе заключалась в предотвращении убытков, что не нравилось многим.

Вдруг он услышал знакомые интонации и впервые поднял глаза на клиента. И заулыбался:

— Василий Иванович! Мы же с вами встречались полтора года назад! Посмотрите, и на визитке вашей только логотип поменялся! А «Системс-Менеджмент» стал «Управлением Системами»!

Клиентщики не понимали, что произошло. Заемщик, похоже, не удивился, но молчал.

— Вы же уже заключали этот контракт, который закончился исками и банкротством...

Этот номер — о профессиональной радости риск-менеджера. К сожалению или к счастью, часто наша радость — результат столкновения с грубыми нарушениями правил, когда хладнокровие и компетентность должны возобладать над эмоциями, особенно в усложняющемся технологическом мире. В этом номере мы попытались распечатать сургуч на таинственном конверте с броским названием AI и посмотреть: как создавать модели, которые не нужно постоянно сопровождать, как защищать их перед регулятором и коллегами, как объяснить их третьим лицам и, что самое важное, — какие вызовы готовит нашей профессии будущее.

С этого вопроса и начнем. И почти одновременно с коллегами из Ruhr University Bochum, которые опубликовали на Github<sup>1</sup> 30-гигабайтную подборку deep fake голосов с претренированными моделями семи типов, отправимся в Дубай, Сингапур и США — вслед за мошенниками, сумевшими обмануть банк синтезированным голосом клиентского топ-менеджера.

Мы также обсудим регулятивные вызовы современности с Алексеем Лобановым, досконально разберемся в LGD-моделях, протестируем новые сервисы проверки контрагентов и побываем на международной финансовой конференции MATLAB вместе с Эмилио Льоренте.

Интересно, что все же это такое — высокотехнологичная радость?

---

<sup>1</sup> <https://github.com/RUB-SysSec/WaveFake>.

---

В базу данных ORX впервые попал крупный инцидент мошенничества с использованием AI. Технология deep voice позволила злоумышленникам вывести из компании \$35 млн. Какие методы помогут банкам противостоять таким атакам и что должно быть приоритетной задачей при защите от них?

## Первая крупная атака с использованием deep fake в базе ORX: что дальше?

Пятнадцатого января 2020 г. Рёдзо Такаги, управляющему директору «дочки» компаний Idemitsu Kosan Co Ltd, позвонили якобы из штаб-квартиры Idemitsu (далее — Idemitsu HQ)<sup>1</sup>. Г-н Такаги идентифицировал голос звонившего как голос Шуничи Кито — президента, директора-представителя и главного исполнительного директора Idemitsu HQ. Звонивший сообщил, что Idemitsu HQ планирует приобретение другой компании, и поручил г-ну Такаги работать с неким Мартином Зельнером (юристом, который якобы сопровождал сделку) для совершения покупки. Затем мошенник попросил г-на Такаги отправить ему электронное письмо, чтобы запросить дополнительную информацию о платеже. В тот же день мошенник, выдававший себя за г-на Зельнера, отправил г-ну Такаги документ с подписью якобы г-на Кито. Этим документом г-ну Такаги были предоставлены полномочия действовать от имени Idemitsu HQ во всем, что касается оплаты, включая подписание документов, работу с банками и иное.

Затем г-ну Такаги были даны указания перевести \$2 978 000 на банковский счет одного лица (далее — D1) и \$2 990 000 на банковский счет другого лица (далее — D2) в Bank of China (Hong Kong) Limited.

Далее 17 и 20 января 2020 г. мошенник, выдавая себя за г-на Кито (по телефону) и г-на Зельнера (через аккаунт своей электронной почты), поручил г-ну Такаги совершить следующие платежи. Г-н Такаги перевел на счет D2 17 января \$3 977 000, 20 января \$2 226 000 и \$3 978 000.

---

<sup>1</sup> Обстоятельства дела изложены по: <https://vlex.hk/vid/idemitsu-chemicals-hong-kong-866704185> и <https://www.documentcloud.org/documents/21085009-hackers-use-deep-voice-tech-in-400k-theft>.

---

## Первая крупная атака с использованием deep fake в базе ORX: что дальше?

---

В материалах дела, выделенного в арбитражный процесс из уголовного, отсутствует подробное описание взаимодействия мошенников со службами банка, обслуживающего платежи клиента, однако указанные крупные переводы по выданной доверенности были совершены.

Третьего февраля 2020 г. мошенник, выдавая себя за г-на Кито, сообщил г-ну Такаги по телефону, что Idemitsu HQ отправит его компании \$100 000 000, и поручил договориться с банком о предоставлении овердрафта. Затем мошенник, выдавая себя за г-на Зельнера, использовал свою учетную запись электронной почты для отправки документа под названием Fund Transfer/Ryozo Takagi, якобы подписанного г-ном Кито (поддельного заявления). В документе говорилось, что Idemitsu HQ переведет \$100 000 000 6 февраля 2020 г.

Очевидно, мошенник (или, скорее всего, высокотехнологичная группа мошенников) после первого успеха решил попробовать «сорвать куш», но с кредитами, в отличие от банковского процесинга, дела обстоят немного сложнее.

После того как г-н Такаги представил поддельную заявку на кредит в банк, она была направлена в Idemitsu HQ, которая раскрыла обман. Idemitsu HQ подтвердила, что г-н Кито не подписывал поддельное заявление, не занимался сделками слияния и поглощения и не давал никаких указаний г-ну Такаги в течение последних нескольких недель.

Пятого февраля 2020 г. об этом деле было сообщено в полицию Гонконга, а 6 февраля 2020 г. дочерняя компания начала судебное разбирательство против D1 и D2. Выяснилось, что деньги были переведены на несколько счетов в других странах в рамках сложной схемы с участием по меньшей мере 17 юридических лиц.

В целом число ответчиков превысило 20. Расследование велось в Гонконге, ОАЭ, США. В ходе расследования, проведенного прокуратурой Дубая, выяснилось, что мошенники использовали технологию deep voice clone. Власти ОАЭ проследили движение денег через многочисленные счета и выявили две транзакции в Соединенные Штаты: 22 января 2020 г. переводы на сумму \$199 987,75 и \$215 985,75 были отправлены в Centennial Bank. Министерство юстиции ОАЭ обратилось к властям США с запросом о судебной помощи, в соответствии с которым был назначен уполномоченный по сбору доказательств от Centennial Bank<sup>1</sup>. В базе ORX событие

---

<sup>1</sup> <https://www.offshorealert.com/united-arab-emirates-centennial-bank/>; <https://www.documentcloud.org/documents/21085009-hackers-use-deep-voice-tech-in-400k-theft>.

отнесено к инцидентам коммерческого банкинга и классифицировано как внешние кражи и мошенничество (тип EL0201)<sup>1</sup>, дословно: «менеджер банка одобрил перечисление 35 млн USD со счета клиента в результате мошенничества с использованием клонирования голоса методом deep fake».

### От редакции

Можно предположить, что мошенником был инсайдер или специалист по социальной инженерии, который немного знал всех сотрудников и сделал «сценарный» deep fake, продумав все возможные вопросы и подставляя ранее сгенерированные ответы. Это лучше, чем потоковая генерация, а известная исполнительность японцев восполнила пробелы в смысловых конструкциях.

## КОММЕНТАРИЙ



Александр  
ДЬЯКОНОВ,  
профессор ВМК МГУ,  
д.ф.-м.н.

Есть два способа подделки голоса:

1. Мультиспикер — система синтеза голоса, то есть перевода text2speech с возможностью синтеза голоса заданным спикером. Для ее обучения нужны как минимум несколько десятков часов транскрибированных аудиозаписей разных спикеров.

Требования к датасетам для синтеза, в отличие от распознавания речи, довольно высокие — записи должны быть чистыми, без шумов. Но все это есть в интернете, кроме того, несложно самому организовать запись датасета с помощью краудсорсинга.

Дальше нужна запись голоса целевого спикера.

Если таких записей достаточно много, есть неплохая возможность fine-tuning — систему можно настроить под нужный голос. Если мало, то может хватить и одной записи. По ней вычисляется «представление спикера». Поскольку система обучалась создавать голос по представлению и тексту, этого может хватить для имитации нужного голоса. Однако использование представлений обеспечивает худшее качество, чем fine-tuning.

2. Конверсия.

Недостаток первого способа в том, что текст переводится в речь. Для имитации живого диалога должен быть чат-бот, который генерирует

<sup>1</sup> <https://managingrisktogether.orx.org/orx-news/5-largest-operational-risk-losses-october-2021>.

операционный риск \ deep voice \ состязательные атаки

---

## Первая крупная атака с использованием deep fake в базе ORX: что дальше?

---

С учетом бурного развития биометрии (например, сгенерированные лица уже почти неотличимы от реальных) банкам стоит быть настороже. Системы, которые анализируют данные, должны учить подобные сценарии и уметь защищаться от них. Далее — о технических аспектах такой защиты.

### Состязательные атаки и механизмы защиты от них

Описанный кейс связан с обманом человека нейронной сетью. Но сама нейронная сеть также должна быть устойчива к атакам мошенников. И здесь возникает проблема состязательных атак<sup>1</sup>. Такие механизмы защиты, как:

- обучение нейронных сетей, устойчивых к состязательным атакам, с использованием как естественных данных обучающей выборки, так и специальным образом зашумленных<sup>2</sup>;
  - дистилляция нейронных сетей<sup>3</sup>;
- 

в режиме реального времени нужные ответы, или возможные сценарии диалога должны быть прописаны (так, например, делают в автоматизированных колл-центрах). Если же чат-бота/сценария нет, да и вообще если диалог может зайти непонятно куда, то проще говорить самому, а некоторая система будет конвертировать голос в целевой. Только вот системы конверсии пока не так совершенны, как мультиспикер.

Прямо сейчас стали появляться очень неплохие на слух решения, но они не позволяют осуществлять конверсию в режиме онлайн. В остальных есть характерные шумы в получаемых записях. Как правило, собеседник может что-то заподозрить, если качество связи хорошее, а если шумы — обычная вещь, то, возможно, речь будет звучать аутентично. В любом случае история кажется довольно неожиданной. Пока что такие технологии не применяются в полной мере — например, в индустрии кино, несмотря на большие бюджеты и множество актуальных задач по озвучке. А здесь мошенники использовали все возможности технологий.

---

<sup>1</sup> [https://github.com/Dyakonov/DL/blob/master/ESSE\\_2021/E036\\_advversarial.pdf](https://github.com/Dyakonov/DL/blob/master/ESSE_2021/E036_advversarial.pdf).

<sup>2</sup> Kurakin A., Goodfellow I., Bengio S. Adversarial Machine Learning at Scale. arXiv:1611.01236, 2016.

<sup>3</sup> Hinton G., Vinyals O., Dean J. Distilling the Knowledge in a Neural Network. arXiv:1503.02531, 2015; Papernot N. et al. Distillation as a Defense to Adversarial Perturbations Against Deep Neural Networks. IEEE Symposium on Security and Privacy (SP), 2016. P. 582–597; Papernot N., McDaniel P. Extending Defensive Distillation. arXiv:1705.05264, 2017.

---

— механизм, основанный на архитектурах GAN<sup>1</sup>, эффективны только в борьбе с конкретным классом атак. Более того, часто из-за использования этих механизмов качество работы сети значительно ухудшается.

Главная причина неудач — свойство transferability модифицированных объектов: модифицированный объект, полученный для атаки одной модели, может быть использован для такой же атаки другой модели независимо от их архитектуры<sup>2</sup>.

Следовательно, ограничение transferability является приоритетной задачей при защите от атак.

## Добавление класса NULL

Один из наиболее действенных способов борьбы со свойством transferability: добавить к выходным классам модели еще один класс — NULL-класс — и обучить классификатор определять модифицированные объекты как объекты из нового класса<sup>3</sup>. Основная идея в том, что при классификации модифицированного объекта модель без такой защиты вынуждена снижать вероятность истинного класса объекта и повышать вероятность других первоначальных классов, тогда как с предложенным расширением у модели появляется возможность повышать вероятность класса NULL без повышения вероятностей других классов. Тем самым модель может отлавливать атаки и отказываться от построения прогноза.

На рис. 1 показан упрощенный пример добавления класса NULL. Чем больше величина модификации, тем большую вероятность классификатор отдает классу NULL.

## Первоначальное обучение

В начале обучения параметры нейронной сети достаточно случайны. Модифицированные объекты, сгенерированные на данном этапе, будут иметь мало общего с объектами, сгенерированными итоговым классификатором. Поэтому изначально модель тренируется на исходных чистых данных. Желаемый выход нейронной сети — это гладкое распределение вероятности принадлежности объектов к классам задачи. То есть желаемая вероятность истинного класса полагается

---

<sup>1</sup> Samangouei P., Kabkab M., Chellappa R. Defense-GAN: Protecting Classifiers Against Adversarial Attacks Using Generative Models. arXiv:1805.06605, 2018.

<sup>2</sup> Papernot N., McDaniel P., Goodfellow I. Transferability in Machine Learning: from Phenomena to Black-Box Attacks using Adversarial Samples. arXiv:1605.07277, 2016.

<sup>3</sup> Hosseini H. et al. Blocking Transferability of Adversarial Examples in Black-Box Learning Systems. arXiv:1703.04318, 2017.

## Первая крупная атака с использованием deep fake в базе ORX: что дальше?

Рисунок 1

### Пример первоначальной разметки и разметки с классом NULL для объектов из датасета MNIST и три объекта для атаки с различными модификациями

Labels	<table border="1"><tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>NULL</td></tr></table>	0	1	2	3	4	5	6	7	8	9	NULL	
0	1	2	3	4	5	6	7	8	9	NULL			
Original Labeling	<table border="1"><tr><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	1	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0			
NULL Labeling	<table border="1"><tr><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	1	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0			
Original Labeling	<table border="1"><tr><td>0.7</td><td>0</td><td>0.3</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	0.7	0	0.3	0	0	0	0	0	0	0	0	
0.7	0	0.3	0	0	0	0	0	0	0	0			
NULL Labeling	<table border="1"><tr><td>0.7</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0.3</td></tr></table>	0.7	0	0	0	0	0	0	0	0	0	0.3	
0.7	0	0	0	0	0	0	0	0	0	0.3			
Original Labeling	<table border="1"><tr><td>0.5</td><td>0</td><td>0.5</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	0.5	0	0.5	0	0	0	0	0	0	0	0	
0.5	0	0.5	0	0	0	0	0	0	0	0			
NULL Labeling	<table border="1"><tr><td>0.5</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0.5</td></tr></table>	0.5	0	0	0	0	0	0	0	0	0	0.5	
0.5	0	0	0	0	0	0	0	0	0	0.5			
Original Labeling	<table border="1"><tr><td>0.1</td><td>0</td><td>0.9</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	0.1	0	0.9	0	0	0	0	0	0	0	0	
0.1	0	0.9	0	0	0	0	0	0	0	0			
NULL Labeling	<table border="1"><tr><td>0.1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0.9</td></tr></table>	0.1	0	0	0	0	0	0	0	0	0	0.9	
0.1	0	0	0	0	0	0	0	0	0	0.9			

равной некоторому числу  $q$ , близкому к единице, вероятности остальных классов распределяются равномерно и равны  $\frac{1-q}{K-1}$ , где  $K$  — число классов в задаче, а вероятность  $p_{\text{NULL}}$  равна нулю.

#### Расчет вероятностей принадлежности к классу NULL для модифицированных объектов

После того как модель достигает высокого качества работы на чистых данных, обучение переходит в следующую фазу. Теперь для каждого обучающего объекта необходимо рассчитать вероятность, что он модифицирован для атаки на нейронную сеть.

Для этого строится отображение  $f:[0,1] \rightarrow [0,1]$ ,  $f(\varepsilon) = p_{\text{NULL}}$  следующим образом.

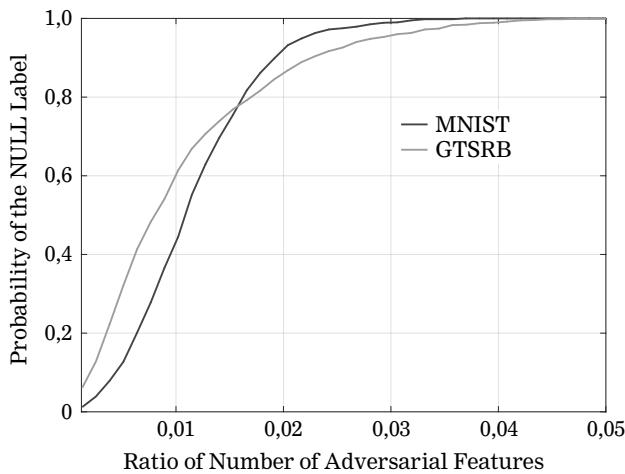
Перебираются объекты из валидационной выборки. Из каждого объекта с помощью методов создания объектов для атак перебором по дискретной сетке значений  $\varepsilon \in [0,1]$  генерируется некоторое подмножество множества  $X_* = \{X_{*,\varepsilon} = X + \delta X_\varepsilon; \|\delta X_\varepsilon\|_0 \leq \varepsilon |X|, \forall \varepsilon \in [0,1]\}$ , где  $\|\cdot\|_0$  — нулевая мера, равная числу ненулевых компонент своего аргумента;  $|X|$  — размерность признакового описания входного объекта  $X$ .

После этого значение функции  $f(\varepsilon)$  полагается равным доле объектов  $X_{*,\varepsilon}$  среди объектов расширенной валидационной выборки, которые оказались успешными в ходе атаки на сеть.

На рис. 2 показаны примеры функции  $f$  для датасетов MNIST и GTSRB.

Рисунок 2

### Вероятность класса NULL для модифицированных объектов в зависимости от доли $\varepsilon$ модифицированных признаков



#### Обучение с учетом состязательных атак

Далее начинается фаза обучения нейронной сети с использованием модифицированных объектов. Эти объекты генерируются из обучающих объектов с помощью метода STG, который модифицирует в обучающем объекте наперед заданное число компонент. Это число выбирается случайным образом из равномерного распределения  $U[1, N_{\max}]$ , где  $N_{\max} \in [1, |X|]$  — это минимальное число, для которого  $f\left(\frac{N_{\max}}{|X|}\right) = 1$ . В результате получается объект  $X_* = X + \delta X$ , желаемое распределение вероятностей которого рассчитывается следующим образом.

Вначале для объекта  $X_*$  рассчитывается вероятность  $p_{\text{NULL}} = f\left(\frac{\|\delta X\|_0}{|X|}\right)$ , затем изначальная вероятность истинного класса  $q$  полагается равной  $q' = q(1 - p_{\text{NULL}})$ , а вероятности остальных классов распределяются равномерно и равны  $\frac{(1-q)(1-p_{\text{NULL}})}{K-1}$ , где  $K$  — число классов в изначальной задаче классификации.

операционный риск \ deep voice \ состязательные атаки

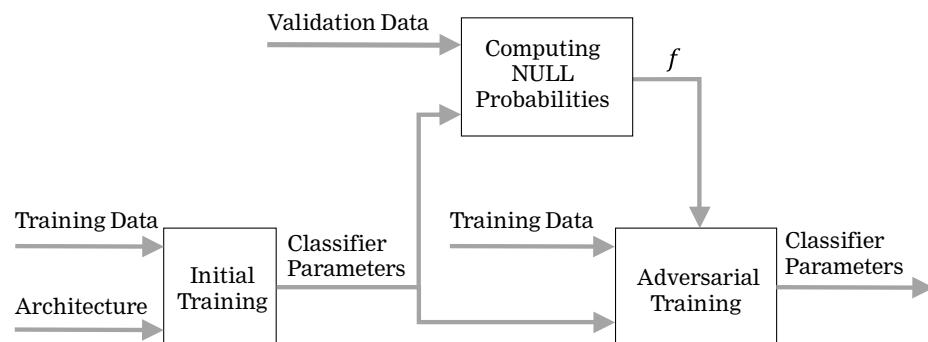
## Первая крупная атака с использованием deep fake в базе ORX: что дальше?

Таким образом, нейронная сеть учится повышать не вероятности неверных классов в случае неуверенности классификации, а вероятность  $p_{\text{NULL}}$  того, что объект является атакующим.

На рис. 3 показана блок-схема работы метода. На фазе начального обучения классификатор обучается только на чистых данных, на фазе обучения с учетом состязательных атак классификатор использует как чистые, так и модифицированные объекты.

Рисунок 3

### Блок-схема метода добавления класса NULL в разметку данных



Этот метод — один из наиболее эффективных в борьбе с состязательными атаками. Он достаточно точно отклоняет атакующие объекты и при этом сохраняет высокое качество работы на естественных объектах.

### Защитные модели на основе GAN-архитектур

Идея алгоритма Defense-GAN<sup>1</sup> — использовать для классификации не поданный на вход классификатора объект  $X$ , а его реконструкцию, полученную с помощью генератора предобученного GAN. Гипотеза такого подхода в том, что если генератор хорошо обучен и достаточно сложен, реконструкция исходного объекта  $X$  будет не сильно отличаться от объекта, но не будет содержать шум  $\delta X$ , который мог быть добавлен в объект  $X$  для атаки на классификатор.

Вначале тренируется GAN  $G(z)$  на имеющихся чистых данных. Затем становится возможным строить с помощью  $G(z)$  реконструкцию

<sup>1</sup> Samangouei P., Kabkab M., Chellappa R. Defense-GAN: Protecting Classifiers Against Adversarial Attacks Using Generative Models. arXiv:1805.06605, 2018.

произвольного входного объекта  $X$ . Для этого решается задача оптимизации

$$\min_z \|G(z) - X\|_2^2$$

и строится подходящая инициализация  $z_*$  для GAN. Теперь можно обучать классификатор, используя чистые данные, реконструкцию чистых данных или совмещая оба подхода.

Когда GAN  $G(\cdot)$  имеет достаточную сложность и процедура обучения хорошо подобрана, имеет место предельное соотношение:

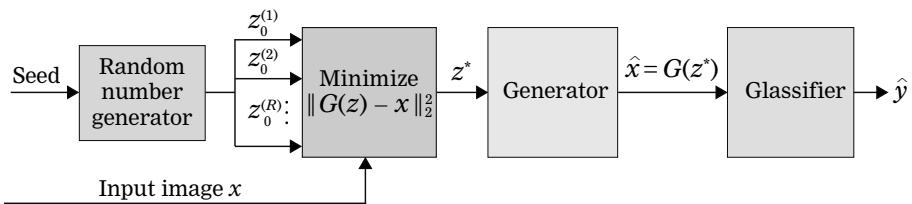
$$E_{x \sim p_{\text{data}}} \left[ \min_z \|G_t(z) - X\|_2 \right] \xrightarrow{t \rightarrow \infty} 0,$$

где  $t$  — шаг обучения GAN<sup>1</sup>.

Так что использование предобученного генератора не должно негативно влиять на процесс обучения классификатора. Схема алгоритма показана на рис. 4. Для классификации входного объекта  $X$  используется его наилучшая реконструкция, построенная с помощью GAN.

Рисунок 4

#### Схема работы алгоритма Defense-GAN



У Defense-GAN есть несколько серьезных преимуществ, выделяющих его среди других средств защиты:

— Defense-GAN можно использовать с любым классификатором. Он никак не изменяет архитектуру классификатора, и его можно считать за шаг предобработки входных данных;

— Defense-GAN способен защищать от любого вида атак, так как его архитектура опирается не на конкретный метод генерации объекта для атаки, а на удаление шума из входного объекта и его сглаживание;

<sup>1</sup> Kabkab M., Samangouei P., Chellappa R. Task-aware compressed sensing with generative models. AAAI Conference on Artificial Intelligence, 2018.

операционный риск \ deep voice \ состязательные атаки

---

## Первая крупная атака с использованием deep fake в базе ORX: что дальше?

---

— Defense-GAN — очень сложная модель, которая итеративно решает задачу оптимизации (для получения оптимальной инициализации  $z_*$ ). Поэтому даже в случае white-box атаки обычные градиентные методы не смогут справиться с задачей построения модифицированного объекта.

Несмотря на то что Defense-GAN оказался эффективным средством защиты от состязательных атак, успех его применения сильно зависит от качества работы генератора GAN. Известно, что хорошее обучение GAN может оказаться трудной процедурой, а без качественного генератора Defense-GAN практически перестает работать. 

Когда Алексей Лобанов был назначен директором Департамента банковского регулирования, мы провели с ним большое интервью, где узнали о планах и перспективах<sup>1</sup>. Спустя 10 лет с того момента, как Алексей Лобанов начал карьеру в Банке России, и через 5 лет после назначения директором Департамента банковского регулирования мы задали ему вопросы наших коллег-рисковиков: что удалось сделать за эти годы, почему регулятор не использует транзакционные данные для регулятивных моделей, а главное, почему требования Банка России столь тяжелы для восприятия и исполнения?

## Алексей Лобанов: итоги 10 лет внедрения Базеля



Алексей ЛОБАНОВ,  
экс-директор Депар-  
тамента банковского  
регулирования Банка  
России

— За 10 лет вам удалось кардинально изменить регулирование. Как бы вы охарактеризовали эти изменения?

— Десять лет назад Россия относилась к тем немногим странам — членам Базельского комитета, где не был внедрен ни один из продвинутых подходов еще предыдущего Базеля — Базеля II. Одной из главных наших задач было быстро создать условия для внедрения хотя бы самого значимого для банков подхода — подхода к оценке кредитного риска на основе внутренних рейтингов.

### От редакции

Согласно табл. 34 статистического бюллетеня Банка России № 230 за декабрь 2021 г., динамика активов кредитных организаций, взвешенных по уровню риска в целях расчета показателя достаточности собственных средств (капитала) Н1.0 в рамках подхода на основе внутренних рейтингов (ПВР), стабильно держится на уровне 25%. По информации редакции, к III кварталу 2021 г. на ПВР перешли три банка из 335.

<sup>1</sup> Банковское регулирование в 2017 году: к чему готовиться риск-менеджерам? // Риск-менеджмент в кредитной организации. 2017. № 1.

## Алексей Лобанов: итоги 10 лет внедрения Базеля

Ситуация усугублялась тем, что у Банка России не хватало полномочий, чтобы, например, внедрять продвинутые подходы, основанные на добровольном выборе банка, требовать соблюдения минимально установленных внутренних процедур оценки достаточности капитала и управления рисками.

Меньше чем через год после начала работы Департамента банковского регулирования вышло Положение № 395-П<sup>1</sup>, которое уже содержало измененное согласно базельским стандартам определение капитала.

### От редакции

В 2012 г. в России действовало более 900 банков, сдававших отчетность по капиталу. Согласно исследованиям ОЭСР по вопросам малого и среднего бизнеса и предпринимательства «Россия: ключевые вопросы и стратегии»<sup>2</sup>, составленным в 2014 г. по запросу Внешэкономбанка, государственного банка развития РФ и Минэкономразвития России, «два важных приоритета заключаются в увеличении плотности размещения банков и повышении правовой защиты для инвесторов. Что касается первого, в Российской Федерации было всего 1112 зарегистрированных банков в 2011 г...., и только 0,169 банковских офисов на 1000 жителей в 2008 г. по сравнению с глобальным средним показателем 0,346 (ОЭСР, 2009). Учитывая географическую концентрацию банковских отделений, особенно в Москве, банковское покрытие является ограниченным в некоторых регионах... Эти проблемы, вероятно, будут усугубляться... у банков, стремящихся... принять новые стандарты “Базель III”».

За неполные 10 лет мы не только обеспечили соответствие практически всему пакету стандартов и рекомендаций Базеля, но и внедрили наиболее интересные и «выгодные» банкам стандарты раньше, чем предусматривает план-график Базельского комитета.

Россия стала едва ли не первой страной, где внедрен новый стандартизованный подход к оценке кредитного риска (с опережением в несколько лет). Это дало возможность банкам высвободить капитал для кредитования лучшей по кредитному качеству части заемщиков. Причем тогда, когда в этом была острая необходимость для поддержки экономики (2019–2020 гг.).

<sup>1</sup> Положение Банка России от 28.12.2012 № 395-П «О методике определения величины собственных средств (капитала) кредитных организаций («Базель III»)».

<sup>2</sup> [https://vrb.ru/common/upload/files/vrb/br/msp/pr2015\\_r.pdf](https://vrb.ru/common/upload/files/vrb/br/msp/pr2015_r.pdf) (с. 102).

## Алексей ЛОБАНОВ

Мы также были первыми, кто в полном объеме внедрил стандартизованный подход к оценке операционного риска. Здесь мы догоняли тот «караван», который ушел еще в нулевых, — когда крупнейшие зарубежные банки переходили на продвинутые подходы.

Что касается регулирования, не связанного с международными требованиями, то я бы назвал несколько интересных задач.

Во-первых, это пропорциональное регулирование (реформа 2016–2018 гг.) — выделение банков с базовой лицензией. Это, возможно, не первая, но наиболее успешная попытка создать нормативную базу для небольших, в основном региональных банков, часто специализирующихся на обслуживании МСП. На момент создания этой нормативной базы на новую лицензию перешли порядка 100 кредитных организаций — очень много для того времени.

Во-вторых, это стимулирующее регулирование — поддержка финансирования приоритетных сегментов, к которым в разное время относились ипотека, МСБ, проектное финансирование. Нам во многом удалось дать банкам стимул с ними работать. Разными путями: от пересмотра регулирования ипотечного кредитования (кульминацией стало внедрение в 2020 г. стандартизированного подхода к расчету кредитного риска для жилой ипотеки) до развития регулирования секьюритизации. Последнее позволило устраниТЬ многолетние сложности и дало сильный импульс участию банков в секьюритизации — причем не только ипотечной, но и (возможно, в первую очередь) секьюритизации кредитов МСП.

Проектное финансирование в регулировании отражено дважды: как (1) особый объект — фабрика проектного финансирования ВЭБ.РФ и (2) остальные объекты, к которым применяются общие подходы. Мы дали банкам возможность оценивать риски инвестиционных проектов не по схеме для действующего предприятия (финансовое положение и качество обслуживания долга), а по большому количеству критериев, характеризующих кредитоспособность проекта вне этих двух «осей».

Еще хотел бы отметить наш успех (пришедшийся на середину 2010-х гг.), который заключался в переводе крупнейшего банка на расчет кредитного риска на основе внутренних рейтингов. Это стало возможным в том числе благодаря взаимодействию с индустрией — иначе мы вряд ли смогли бы справиться, особенно в столь краткий срок.

Сейчас таких банков три, четвертый проходит валидацию. В совокупности это больше половины активов банковского сектора (к моменту, когда все эти банки выполнят свои планы перехода).

«За неполные 10 лет мы не только обеспечили соответствие практически всему пакету стандартов и рекомендаций Базеля, но и внедрили наиболее интересные и “выгодные” стандарты раньше, чем предусматривает план-график Базельского комитета».

## Алексей Лобанов: итоги 10 лет внедрения Базеля

Банк России видит за этим подходом большое будущее. В 2021 г. мы обсудили с банковским сообществом консультативный доклад<sup>1</sup>, где рассматривали возможность перевода всех системно значимых банков на подход на основе внутренних рейтингов в безальтернативном порядке, чтобы дать возможность равной конкуренции, исключить арбитраж, а главное — привести регулятивные требования в соответствие с тем уровнем управления рисками, которого мы вправе ожидать от таких банков.

— То есть спрос на моделистов возрастет.

— Если эта идея получит развитие, то да.

— Тщательность Банка России в подготовке нормативных документов хорошо известна. Это документы на десятки страниц, с большим количеством приложений, таблиц, агрегаторов, расчетов, схем и т.д. Как готовится эта документация? Вы отталкиваетесь от какой-то модели? Или от опыта? Если мы поймем, как эти документы создаются, то сможем понять, как их исполнять.

— Значительная часть наших нормативных актов представляет собой национальную реализацию международных стандартов и рекомендаций. Это касается большей части обязательных нормативов, определения капитала, нормативов ликвидности. Требования к ВПОДК — это тоже во многом реализация международных стандартов и опыта. Требования к раскрытию информации банками тоже являются по сути международными стандартами. Эти документы появились в разное время, в разных условиях, но их объединяет то, что они реализуют уже существующий текст. Мы вынуждены соблюдать ограничения: например, использовать только те понятия, которые уже существуют в праве.

С другой стороны, наше регулирование исторически сильно зависит от правил бухгалтерского учета. Россия относится к тем государствам, где регулирование основано преимущественно на правилах, а не на принципах, — в отличие от, например, стран англосаксонской системы права. В данной ситуации важно детализировать регулирование до такого уровня, чтобы оно было рабочим. Под этим я понимаю возможность банков применять его самостоятельно, задавая минимум вопросов регулятору. Отсюда и симбиоз регулирования и бухучета. Чтобы можно было корректно рассчитать показатель,

<sup>1</sup> <https://cbr.ru/press/event/?id=10904>.

## Алексей ЛОБАНОВ

агрегат, коэффициент, необходимо указать счета (части счетов), на которых находятся нужные для расчета величины.

Если бы этого не было — как в странах, где регулирование дословно, без детализации, повторяет базельский текст, — были бы большие разнотчтения между банками. А следствие — неспособность надзора увидеть, в чем заключается проблема (например, выявить манипуляции с нормативами).

Поэтому в большинстве нормативных актов, имеющих международную природу, можно увидеть «ядро», описанное русским юридическим языком с соответствующей адаптацией, и внешнюю «обвязку» — ту страновую специфику, о которой нет ни слова в международных документах.

Нормативные документы, российские «по происхождению», например Положение № 590-П, — это рамки для внутренних методик банков. У таких документов долгая интересная судьба: в них вносятся многочисленные изменения, которые утяжеляют текст, делают его трудным для понимания, но отражают, насколько возможно, реакцию банков на предыдущее регулирование.

### От редакции

В неформальном исследовании, проведенном нашим ИД в ноябре 2021 г., Положение № 590-П фигурировало в объявлениях на hh.ru в 1,4 раза чаще, чем Python.

— **Учитывается ли при имплементации международных норм транзакционная активность банков? Ведь Банк России видит каждую проводку, более или менее в реальном времени.**

— При принятии любого нормативного акта оценивается регулирующее воздействие. Мы даем возможность прислать комментарии, в том числе расчеты. «Проиграть» самим, внутри, влияние на некую модель банка мы можем в ограниченном количестве случаев — когда у нас есть все данные из отчетности, которую нам сдают банки. Но отчетность банков спроектирована исходя из иных задач, нежели самостоятельный расчет тех или иных агрегаторов регулятором. Поэтому она часто не содержит той детальной гранулированной информации, которая есть у банка и которую нужно знать для корректного расчета рисков.

— **Что нас ждет с точки зрения регулирования?** Например, на Западе банковские модели усложняются. Регулирование во многом учитывает финтех-тенденции, усложнение и рост

Российское регулирование основано преимущественно на правилах, а не на принципах.

В такой ситуации важно его детализировать до такого уровня, чтобы оно было рабочим.

Отсюда симбиоз регулирования и бухучета.

## Алексей Лобанов: итоги 10 лет внедрения Базеля

**количества данных, Explainable AI, права человека в условиях бурного развития искусственного интеллекта. Насколько это актуально для нас?**

— Мы старались учесть эти тенденции, например в регулировании операционного риска. Финтех очень разный, поэтому интегральное регулирование вряд ли возможно. Скорее это будет регулирование по видам деятельности: если финтех воспроизводит банковскую услугу, он должен подвергаться банковскому регулированию. Полагаю, мы будем двигаться от регулирования по видам лицензий к регулированию по видам деятельности. И кстати, этот вопрос затронут в консультативном докладе «Совмещение видов деятельности на финансовом рынке»<sup>1</sup>. Скорее всего, это работа как минимум на несколько ближайших лет.

### От редакции

Пользуясь случаем, вспомним и мы эти десять лет. Десять лет назад, даже семь лет назад, банковский бизнес все еще был «бизнесом» — его можно было купить, продать, он генерировал прибыль частным инвесторам, существовали известные посредники по покупке региональных банков и сделки M&A «за несколько капиталов» были нередки. Коммерсанты ходили по частным банкам, звонили «своему банкиру» и получали кредиты, многие зарабатывали несколько десятков «иксов» в год. Бизнес развивался, росла розница. Все что-то покупали или производили. Действовал очень известный МБК-форум, и откровения на нем стоили стрессов диалога с регулятором. В чем-то это было похоже на спорт, а по вечерам все сбрасывали стресс и в сущности жили одним днем, держась за своего акционера. Но что-то поменялось в корне, и сейчас банковское дело стало сродни государственному управлению. Закончим статью словами Алексея Анатольевича, с которыми мы согласны:

«И если нам удалось что-то сделать за эти годы не для повышения устойчивости банков, стимулирования кредитования или устранения регуляторных барьеров, а для того, чтобы профессия риск-менеджера в банке получила поддержку, признание и перспективы, — в конечном счете это самое главное. Важно, чтобы в индустрии было достаточно активных коллег, с кем можно вести диалог на одном языке». ■

<sup>1</sup> <https://cbr.ru/press/event/?id=11148>.

---

В предыдущей статье<sup>1</sup> мы разобрали первые три фазы проекта разработки модели LGD согласно стандарту CRISP-DM: понимание целей задачи, изучение данных и подготовка данных. Разработка модели является следующей фазой, которую в рамках методологии ПВР и МСФО (IFRS) 9 можно разделить на три этапа: разработка «ядра» модели, калибровка модели, применение надбавок. Рассмотрим первый этап.

**Сергей АФАНАСЬЕВ**, КБ «Ренессанс Кредит», вице-президент, начальник управления статистического анализа

**Анастасия СМИРНОВА**, КБ «Ренессанс Кредит», начальник отдела разработки и анализа эффективности скоринговых систем

**Ильгиз АХМЕТСАФИН**, КБ «Ренессанс Кредит», руководитель направления разработки рейтинговых систем

**Игорь МОЛОКАНОВ**, КБ «Ренессанс Кредит», главный аналитик направления разработки рейтинговых систем

## LGD-модели для розничного кредитования. Часть 2: разработка «ядра» модели

На рис. 13 показана развернутая схема разработки модели LGD со списком работ на каждом этапе проекта. Более темным цветом выделены этапы, относящиеся к фазе «Разработка модели» согласно циклу CRISP-DM. Процесс разработки «ядра» модели включает в себя несколько этапов анализа, состав и порядок которых могут меняться в зависимости от типа задачи и выбранного подхода к моделированию:

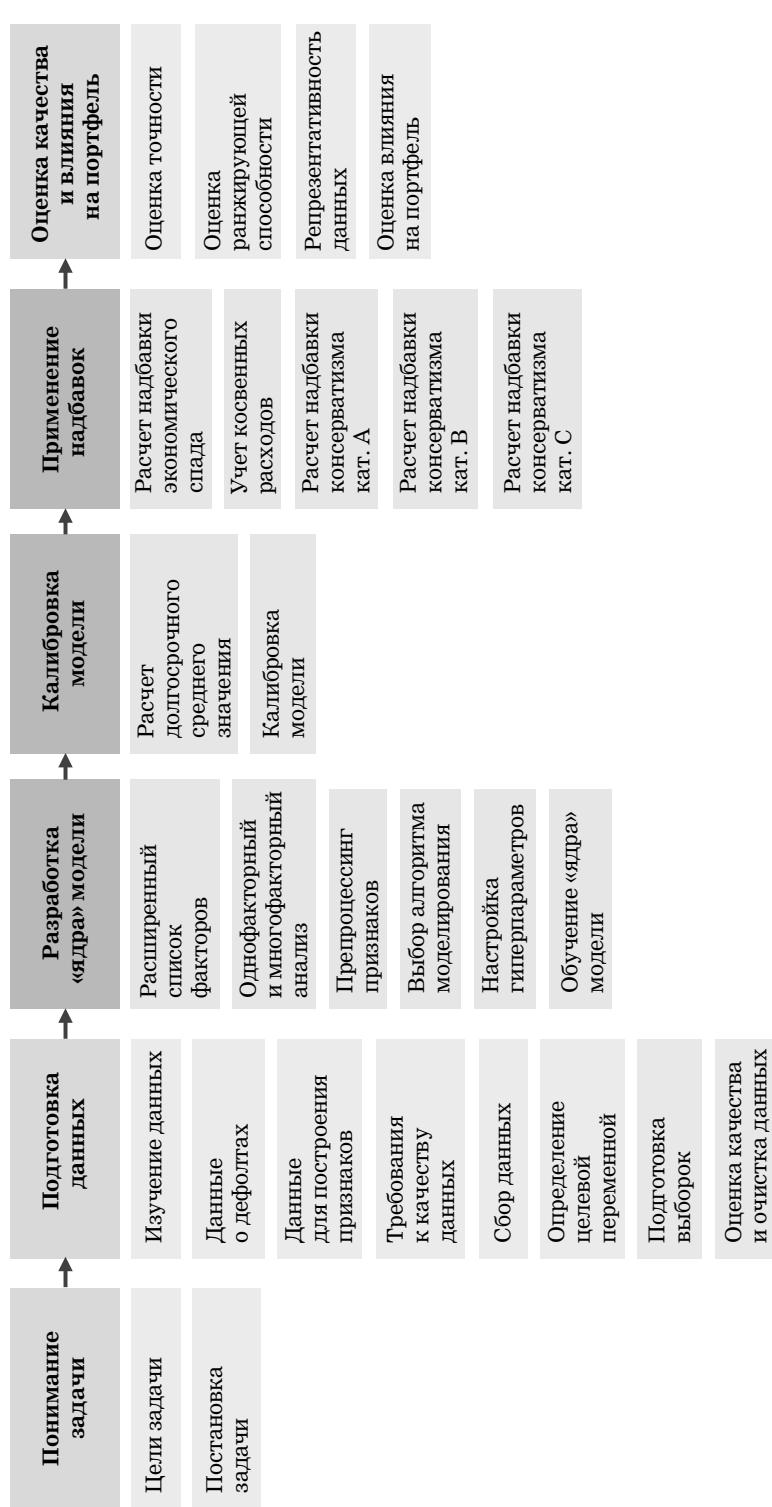
1. *Разработка расширенного списка признаков* — на этом этапе разрабатываются признаки, дополняющие исходный список признаков, подготовленный в процессе формирования выборки. Данный этап не всегда включается в процесс разработки «ядра» модели, а решение о проведении данного этапа может зависеть от структуры исходной выборки, свойств алгоритма моделирования и типа задачи. Например, если в исходной выборке содержится большое количество разработанных экспертами признаков, аналитик может принять

---

<sup>1</sup> Афанасьев С. и др. Разработка LGD-моделей для розничного кредитования. Часть 1: подготовка данных // Риск-менеджмент в кредитной организации. 2021. № 3.

Рисунок 13

### Схема разработки модели LGD со списком работ на каждом этапе



## Сергей АФАНАСЬЕВ и др.

решение не дополнять исходный список, упрощая тем самым дальнейший анализ данных и этап отбора переменных.

2. *Отбор переменных* — на данном этапе проводится однофакторный и многофакторный анализ переменных (признаков). Для классических алгоритмов данный этап необходим, так как расширенный список признаков часто является избыточным и может снижать качество модели за счет эффекта мультиколлинеарности переменных. В продвинутых алгоритмах проблема переобучения моделей может быть решена с помощью различных техник регуляризации, таких как L1 и L2, pruning, dropout, batch normalization и др.

3. *Препроцессинг данных* — на данном этапе проводятся преобразования данных, позволяющие улучшить качество модели. Например, для алгоритмов регрессии важно, чтобы переменные имели распределение, близкое к нормальному. Для этого используются различные техники преобразования переменных.

4. *Выбор алгоритма моделирования* — на данном этапе выбирается тип статистического алгоритма для обучения модели, исходя из свойств задачи, предпочтений аналитика, внутренних политик банка, требований регулятора и мнений экспертов, вовлеченных в процессы моделирования, валидации и применения модели.

5. *Настройка гиперпараметров и обучение модели* — на данном этапе проводится настройка выбранных гиперпараметров модели. Данный этап позволяет улучшить качество модели за счет подбора гиперпараметров, а также автоматизировать часть ручных процессов. Например, этап отбора переменных может быть заменен на подбор коэффициента регуляризации, обнуляющего или понижающего веса избыточных переменных. Решение о включении данного этапа в процесс разработки «ядра» модели принимается аналитиком в зависимости от типа выбранного алгоритма и свойств задачи.

6. *Обучение «ядра» модели* — на данном финальном этапе проводятся обучение «ядра» модели и проверка его качества на тестовой выборке.

Этап отбора признаков мы детально разбирали в одной из статей<sup>1</sup>, поэтому перейдем сразу к этапу препроцессинга данных.

### Препроцессинг данных

После отбора признаков выборку с коротким списком переменных необходимо подготовить для обучения «ядра» модели в соответствии

<sup>1</sup> Афанасьев С., Котерева Д., Смирнова А. Two-forest jupt: комбинированный отбор признаков с использованием двухлесостороннего метода // Риск-менеджмент в кредитной организации. 2020. № 3.

## LGD-модели для розничного кредитования. Часть 2: разработка «ядра» модели

с требованиями выбранного алгоритма и другими эвристиками, позволяющими улучшить качество финальной модели.

Качество линейных алгоритмов, к которым относятся логистическая и линейная регрессия, зависит от некоторых свойств данных. В частности, признаки должны быть нормализованы, то есть иметь одинаковый масштаб. Данное свойство необходимо как для градиентных методов (логистическая регрессия), где нормализация признаков позволяет улучшать качество модели<sup>1</sup>, так и для аналитических (метода наименьших квадратов в линейной регрессии), в которых нормализация признаков позволяет привести коэффициенты регрессии к одинаковому масштабу для дальнейшего анализа вклада переменных в результат.

Среди методов нормализации признаков можно выделить:

1. Стандартизацию для симметрично распределенных признаков:

$$\hat{x}_i = \frac{x_i - \bar{x}_i}{\sigma(x_i)}, \quad (10)$$

где  $\hat{x}_i$  — стандартизованный  $i$ -й признак;

$x_i$  — исходный  $i$ -й признак;

$\bar{x}_i$  — среднее значение  $i$ -го признака;

$\sigma(x_i)$  — среднеквадратичное отклонение  $i$ -го признака.

2. Логарифмическое преобразование для асимметрично распределенных признаков:

$$\hat{x}_i = \log(x_i + \varepsilon), \quad (11)$$

где  $x_i$  — исходный  $i$ -й признак;

$\log$  — логарифм по основанию 10,  $e$  или 2;

$\varepsilon$  — константа, подобранная так, чтобы выражение под логарифмом было больше нуля.

3. Преобразование Бокса–Кокса<sup>2</sup>, позволяющее подбирать оптимальную трансформацию той или иной переменной:

$$\hat{x}_i(\lambda) = \begin{cases} \frac{x_i^\lambda - 1}{\lambda}, & \text{если } \lambda \neq 0; \\ \log(x_i), & \text{если } \lambda = 0, \end{cases} \quad (12)$$

где  $x_i$  — исходный  $i$ -й признак;

$\log$  — логарифм по основанию 10,  $e$  или 2;

$\lambda$  — подбираемый параметр.

<sup>1</sup> Воронцов К.В. Лекции по линейным алгоритмам классификации. 19 января 2009 г. (<http://www.machinelearning.ru/wiki/images/6/68/voron-ML-Lin.pdf>).

<sup>2</sup> Box G.E.P, Cox D.R. (1964). An analysis of transformations // Journal of the Royal Statistical Society, Series B. Vol. 26. P. 211-252.

## Сергей АФАНАСЬЕВ и др.

Подбор  $\lambda$  осуществляется через максимизацию логарифма правдоподобия. При этом значения признаков должны быть *положительными*.

4. Семейство преобразований Йео–Джонсона<sup>1</sup>, которое позволяет обойти ограничения на отрицательные и нулевые значения признаков:

$$\hat{x}_i(\lambda) = \begin{cases} \frac{(x_i + 1)^\lambda - 1}{\lambda}, & \text{если } \lambda \neq 0, x_i \geq 0; \\ \log(x_i + 1), & \text{если } \lambda = 0, x_i \geq 0; \\ \frac{(1-x_i)^{2-\lambda} - 1}{\lambda-2}, & \text{если } \lambda \neq 2, x_i < 0; \\ -\log(1-x_i), & \text{если } \lambda = 2, x_i < 0, \end{cases} \quad (13)$$

где  $x_i$  — исходный  $i$ -й признак;

$\log$  — логарифм по основанию 10,  $e$  или 2;

$\lambda$  — подбираемый параметр.

Подбор параметра  $\lambda$  осуществляется через минимизацию расстояния Кульбака–Лейблера между нормальным и преобразованным распределениями.

## Выбор подхода к моделированию

Оценка уровня потерь при дефолте относится к задачам регрессии, где результатом предсказания является непрерывная величина LGD. С другой стороны, в силу специфики процесса кредитования значения LGD обычно распределяются по бимодальному закону (рис. 14), из-за чего применение классических регрессионных подходов может давать плохие результаты<sup>2</sup>.

Поэтому на практике для прогнозирования LGD могут использоваться как однокомпонентные, так и многокомпонентные регрессионные алгоритмы:

1. Подход Gross — строится регрессионная модель, предсказывающая компоненту LGD.
2. Подход Cure/Loss Rate — строится комбинированная модель с использованием бинарного классификатора и алгоритма регрессии.

<sup>1</sup> Yeo I.-K., Johnson R.A. (2000). A New Family of Power Transformations to Improve Normality or Symmetry // Biometrika. Vol. 87. No. 4. P. 954-959. DOI: 10.1093/biomet/87.4.954. JSTOR 2673623.

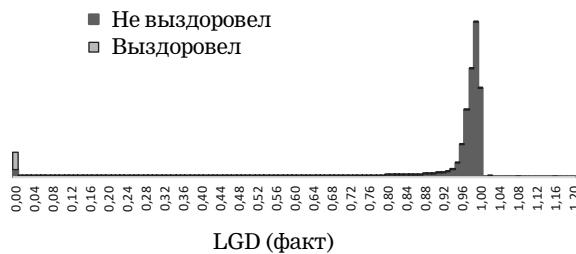
<sup>2</sup> Heckman J. (1979). Sample Selection Bias as a Specification Error // Econometrica. Vol. 47. No. 1. P. 153-161. DOI: 10.2307/1912352. JSTOR 1912352. MR 0518832.

## LGD-модели для розничного кредитования. Часть 2: разработка «ядра» модели

Рисунок 14

### Примеры бимодального распределения cash- и POS-кредитов по значению LGD

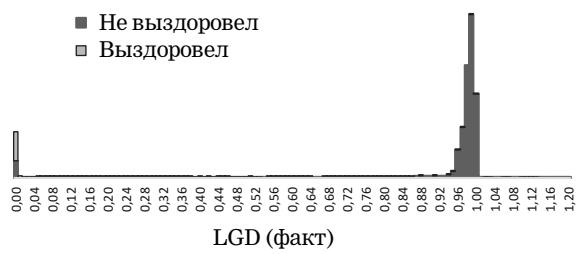
Cash-кредиты. Модель LGD “In-Default”



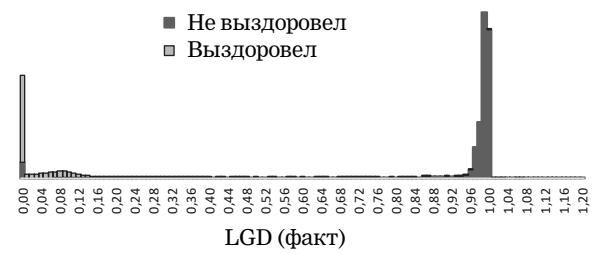
Cash-кредиты. Модель LGD “Will-Default”



POS-кредиты. Модель LGD “In-Default”



POS-кредиты. Модель LGD “Will-Default”



### Подход Gross

В рамках методологии Gross строится одна модель: моделирование LGD производится непосредственно через предсказание целевой переменной<sup>1</sup>. При построении модели значения целевой переменной (историческое значение LGD) ограничиваются на интервале от 0% до 120% для получения консервативных результатов и уменьшения влияния выбросов. Преимуществом данного подхода является его простота. Недостатком метода может быть невысокая точность оценок из-за мультимодального характера распределения кредитов по значениям LGD.

<sup>1</sup> Методика расчета целевой переменной приводилась в первой части статьи в прошлом номере журнала.

## Сергей АФАНАСЬЕВ и др.

### Подход Cure/Loss Rate

В рамках подхода Cure/Loss Rate модельная оценка LGD строится через моделирование нескольких компонент<sup>1</sup>:

$$LGD = CR \times LR_c + (1 - CR) \times LR_t, \quad (14)$$

где  $LR_c$  и  $LR_t$  — уровень потерь при дефолте для «выздоровевших» и «не выздоровевших» кредитов соответственно;

$CR$  — вероятность «выздоровления» кредита.

Для наблюдений из выборки для разработки модели LGD исторические значения  $CR$  могут быть 1 в случае закрытия дефолта со статусом «выздоровление»<sup>2</sup> или 0 в случае закрытия дефолта с прочими статусами. Исторические значения  $LR$  моделируются аналогично подходу Gross и ограничиваются интервалом от 0% до 120% с целью получения консервативных результатов и уменьшения влияния выбросов.

Для моделирования компоненты  $CR$  используются все статистические единицы, попавшие в выборку для разработки модели LGD. Для расчета  $LR_c$  используются только те контракты, по которым событие дефолта было закрыто статусом «выздоровление», а для расчета  $LR_t$  используются контракты, по которым событие дефолта было закрыто с прочими статусами. Таким образом, с точки зрения целевой переменной:

- величина  $CR$  является бинарной (0 или 1);
- величины  $LR_c$  и  $LR_t$  являются непрерывными и ограничиваются снизу 0% для получения более консервативных оценок, а также могут быть ограничены сверху 120% с целью уменьшения влияния выбросов.

Результатом данного подхода может являться матрица, в которой по осям находятся бакеты с интервалами значений  $CR$  и  $LR$  (при использовании методики дерева решений для компонент  $LR$  и разделении непрерывных результатов  $CR$  на группы значений), а их пересечения — модельные значения LGD. При использовании хотя бы одной из компонент модели в непрерывной форме результатом модели является применение формулы (14), в которой производится умножение значений  $LR$  на величины  $CR$  и  $(1 - CR)$ . Преимуществом данного подхода является разделение на большее количество модельных

Если LGD имеет мульти-модальный характер распределения, для построения «ядра» модели рекомендуется использовать подход Cure/Loss Rate — строить комбинированную модель с использованием бинарного классификатора и алгоритма регрессии.

<sup>1</sup> Bellini T. IFRS 9 and CECL Credit Risk Modelling and Validation, гл. 4.

<sup>2</sup> Определение статуса «выздоровление» зависит от бизнес-процесса. В качестве критериев «выздоровления» могут рассматриваться как наступление определенных событий по кредиту, так и определенное значение LGD. К «выздоровевшим» дефолтам относятся в том числе «погашенные» дефолты.

## LGD-модели для розничного кредитования. Часть 2: разработка «ядра» модели

групп на основании различных характеристик, позволяющее получить более точные оценки. К недостаткам подхода можно отнести высокие требования к качеству и полноте данных, требования к наличию процесса «выздоровления» на всем горизонте выборки для разработки, а также определение критериев «выздоровления», которые позволяют выделить моды в распределении LGD.

Если LGD имеет мультимодальный характер распределения, для построения «ядра» модели рекомендуется использовать подход Cure/Loss Rate<sup>1</sup>.

### Алгоритм для разработки «ядра» модели

В зависимости от используемой методологии расчета целевой переменной для построения «ядра» модели LGD оценивается одна или несколько компонент LGD. Для моделирования непрерывных целевых переменных используются различные алгоритмы машинного обучения — от линейной регрессии до глубоких нейронных сетей. Для моделирования компонент моделей LGD (LR<sub>c</sub>, LR<sub>t</sub>, CR) в рамках регуляторных требований ПБР и МСФО (IFRS) 9 на практике применяют следующие алгоритмы:

- 1) look-up таблица;
- 2) линейная регрессия;
- 3) дерево решений;
- 4) логистическая регрессия (для моделирования компоненты CR).

### Look-up таблица

Подход Look-up является наиболее простым методом разработки модели LGD, который применяют в случае недостаточного количества наблюдений в выборке для разработки. Данный подход может обеспечивать более устойчивые результаты моделирования, когда результаты других методов могут быть неудовлетворительными.

Look-up таблица может рассматриваться как дерево решений, в котором правила разделения значений признаков не создаются автоматическим алгоритмом, а выполняются аналитиком посредством одномерного и многомерного анализа. После завершения кластеризации значений количественных и категориальных признаков для каждого полученного бакета рассчитывается статистика по целевой переменной (среднее значение или медиана).

Look-up таблица строится по следующему алгоритму (табл. 1):

<sup>1</sup> В некоторых видах кредитования со сложной структурой залогов распределение кредитов по LGD может содержать больше двух локальных мод, поэтому для таких портфелей строятся мультиклассовые модели.

## Сергей АФАНАСЬЕВ и др.

Таблица 1

### Пример разработки модели LGD с помощью Look-up таблицы

	Группа 1	Группа 2	Группа 3
Группа 1	LGD <sub>1</sub>	LGD <sub>2</sub>	LGD <sub>3</sub>
Группа 2	LGD <sub>4</sub>	LGD <sub>5</sub>	LGD <sub>6</sub>
Группа 3	LGD <sub>7</sub>	LGD <sub>8</sub>	LGD <sub>9</sub>
Группа 4	LGD <sub>10</sub>	LGD <sub>11</sub>	LGD <sub>12</sub>

1. Первый признак с самым высоким уровнем значимости используется для дифференциации наблюдений между бакетами.
2. Второй признак с высоким уровнем значимости включается в анализ для дифференциации риска, а результаты переоцениваются.
3. Те же шаги проделываются для всех остальных признаков, включенных в анализ.

При использовании Look-up подхода бакеты формируются таким образом, чтобы наиболее однородные наблюдения были сгруппированы. При необходимости некоторые бакеты объединяются, если удается обнаружить схожие характеристики у наблюдений внутри рассматриваемых групп. Данный процесс применяется ко всем признакам, прошедшим отбор через однофакторный и многофакторный анализ.

К преимуществам подхода Look-up относят стабильность прогнозов на маленьких выборках и простоту использования. Среди минусов выделяют невысокую точность прогнозов и зависимость результатов от мнения экспертов.

### Линейная регрессия

В рамках регрессионного анализа используется модель множественной регрессии для оценки прогнозных значений LGD<sup>1</sup>:

$$Y = \sum_i \beta_i \times X_i + \varepsilon \quad (15)$$

где  $Y$  — зависимая переменная;

$\beta_i$  — коэффициенты регрессии;

$X_i$  — объясняющие переменные (признаки);

$\varepsilon$  — случайная ошибка модели.

Для получения хороших, устойчивых результатов линейной регрессии важно корректно оценивать регрессионные коэффициенты,

<sup>1</sup> Engelmann B., Rauhmeier R. Basel II Risk Parameters, гл. 11.5.3.

## LGD-модели для розничного кредитования. Часть 2: разработка «ядра» модели

которые должны быть несмещеными. Для этого необходимо соблюдение следующих условий:

- правильный выбор спецификации модели, подтверждаемый линейной зависимостью между независимыми переменными и зависимой переменной;
- отсутствие сильной линейной зависимости между признаками (низкая мультиколлинеарность);
- низкая корреляция между случайной ошибкой и независимыми переменными (слабая экзогенность признаков);
- постоянство дисперсии случайных ошибок (гомоскедастичность);
- нормальное распределение ошибок.

Среди плюсов линейной регрессии можно выделить хорошую интерпретируемость результатов, непрерывность модельных значений и низкую вероятность переобучения относительно других подходов. К недостаткам метода относят необходимость выполнения большого количества требований к данным и выход модельных оценок за пределы логически допустимого интервала значений.

### Дерево решений

Один из наиболее популярных алгоритмов CART предназначен для построения бинарного дерева решений<sup>1</sup>, в котором каждый узел при разбиении имеет только двух потомков. При построении дерева на каждом шаге формируемое в узле правило делит заданное множество наблюдений на две части: (1) часть, в которой правило выполняется (потомок “right”); (2) часть, в которой правило не выполняется (потомок “left”). Начиная с полной выборки наблюдения перераспределяются по узлам дерева в зависимости от значений признаков, при этом подвыборки узлов более глубокого уровня становятся более однородными по отношению к целевой переменной (рис. 15).

Плюсы алгоритма дерева решений:

- по сравнению с линейной регрессией дерево решений не требует выполнения большого количества предпосылок для применения;
- пропущенные значения не требуют специальной обработки, поскольку дерево решений способно выделить их в отдельную категорию;
- мультиколлинеарность среди независимых переменных не является ограничением для построения дерева решений;

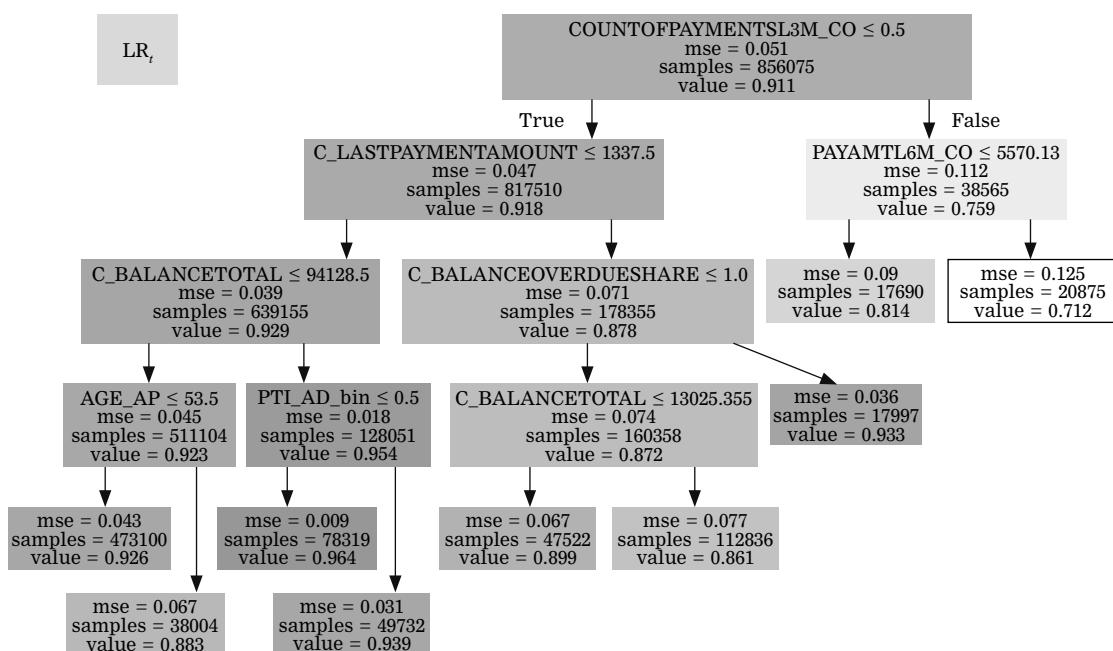
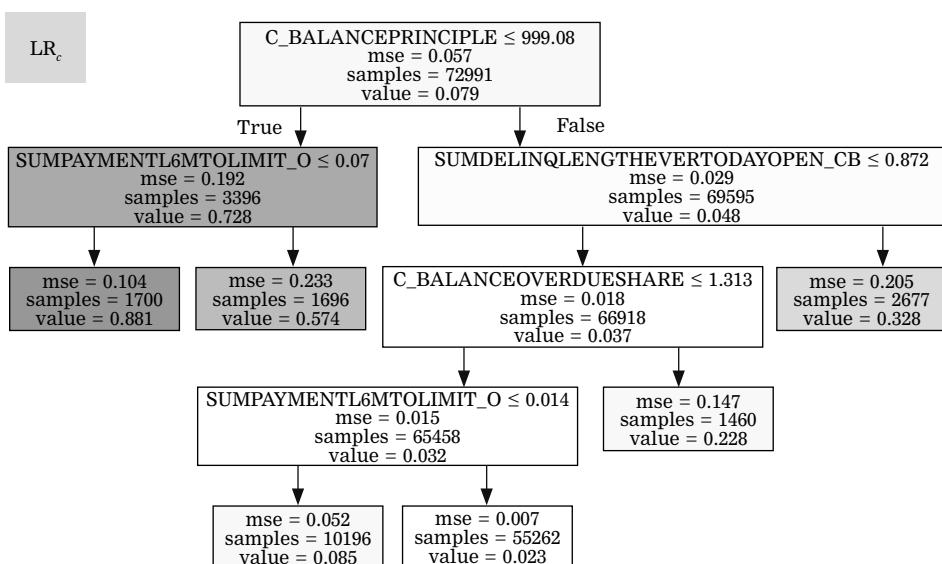
Для получения хороших, устойчивых результатов линейной регрессии важно корректно оценивать регрессионные коэффициенты, которые должны быть несмещеными.

<sup>1</sup> Breiman L., Friedman J.H., Olshen R.A., Stone C.J. (1984). Classification and regression trees. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software.

**Сергей АФАНАСЬЕВ и др.**

Рисунок 15

**Пример модели LGD, построенной на дереве решений (подход Cure/Loss Rate)**



## LGD-модели для розничного кредитования. Часть 2: разработка «ядра» модели

- дерево решений способно распознать зависимость целевой переменной и признаков даже в случае мультимодального распределения целевой переменной (в отличие от линейной регрессии, где мультидимодальность может стать проблемой);
- алгоритм обеспечивает однородность конечных подвыборок;
- высокая интерпретируемость алгоритма — понятная визуальная структура дерева решений позволяет увидеть логику сегментации.

Среди минусов алгоритма выделяют следующие:

- деревья решений могут сильно переобучаться, из-за чего требуется дополнительный контроль параметров дерева;
- деревья решений относятся к «жадным алгоритмам», то есть в процессе обучения модели используются наиболее оптимальные разбиения на каждом узле даже в случаях, когда можно отобрать уже использованный признак без потери в качестве. Таким образом, при обучении дерева решений может быть задействовано большое количество признаков, что повышает требования к качеству данных.

### Логистическая регрессия

В рамках подхода Cure/Loss Rate для оценки параметра CR необходимо использовать алгоритмы классификации, которые позволяют получить вероятность того, что событие дефолта будет закрыто путем «выздоровления».

Для классификации используются логистическая регрессия, деревья решений, метод опорных векторов, случайный лес, градиентный бустинг, нейронные сети и др. Регулярно проводимые сравнения методов машинного обучения (бенчмарки) показывают, что сами по себе алгоритмы бинарной классификации не сильно различаются по качеству, а финальные результаты во многом зависят от умения работать с особенностями того или иного алгоритма<sup>1</sup>.

Наиболее популярным алгоритмом бинарной классификации для разработки моделей ПВР и МСФО (IFRS) 9 является логистическая регрессия, среди достоинств которой можно выделить хорошую интерпретируемость и робастность (устойчивость к «шуму»). При использовании логистической регрессии важно, чтобы доступные признаки позволяли с высокой точностью предсказывать («разделять») классы объясняемой переменной.

Дерево решений способно распознать зависимость целевой переменной и признаков даже в случае мультимодального распределения целевой переменной (в отличие от линейной регрессии, где мультидимодальность может стать проблемой).

<sup>1</sup> Munkhdalai L. et al. (2019). An Empirical Comparison of Machine-Learning Methods on Bank Client Credit Assessments // Sustainability | An Open Access Journal from MDPI. Vol. 11. Issue 3. P. 1-23. DOI: 10.3390/su11030699; Lessmann S. et al. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. DOI: 10.1016/j.ejor.2015.05.030.

---

Сергей АФАНАСЬЕВ и др.

---

## Гиперпараметры «ядра» модели

Алгоритмы машинного обучения, включая деревья решений и логистическую регрессию, имеют большое количество настроек — гиперпараметров модели. Значения гиперпараметров устанавливаются перед запуском процесса обучения модели. В этом смысле гиперпараметры отличаются от обычных параметров, настраиваемых в процессе обучения модели.

В деревьях решений наиболее распространеными гиперпараметрами, влияющими на качество модели, являются глубина дерева, минимальное количество наблюдений в листе и минимальное количество наблюдений, необходимое для разделения в узле. При большой глубине дерева или маленьком количестве наблюдений в листах алгоритм может переобучиться, то есть снизится обобщающая способность модели. В связи с этим в процессе обучения дерева решений рекомендуется настраивать указанные гиперпараметры.

В многомерной линейной регрессии и логистической регрессии наиболее распространеными гиперпараметрами являются коэффициент регуляризации и коэффициент балансировки классов (для задачи классификации), которые необходимо подбирать перед началом обучения модели.

Настройка гиперпараметров модели осуществляется с помощью метода перекрестной проверки (cross-validation)<sup>1</sup>. Основная идея метода в том, что каждое наблюдение выборки может быть использовано для проверки. В машинном обучении существуют различные методы перекрестной проверки, среди которых наиболее распространена  $k$ -блочная перекрестная проверка ( $k$ -fold cross-validation). Метод  $k$ -блочной перекрестной проверки как частный случай общего метода состоит в том, что выборка разбивается на  $k$  блоков, где один из блоков используется для проверки, а остальные  $k - 1$  блоков объединяются в обучающую выборку для построения модели (рис. 16). В результате  $k$ -блочной перекрестной проверки получается  $k$  разных обученных моделей. Обобщающая способность модели, оцениваемая выбранной метрикой качества, вычисляется как среднее арифметическое  $k$  оценок, полученных на валидационных блоках.

Гиперпараметры модели подбираются поиском по сетке значений (Grid Search), заданных аналитиком, путем обучения модели с помощью метода  $k$ -блочной перекрестной проверки. При этом для подбора гиперпараметров используется только обучающая выборка, а тестовая

---

После обучения «ядра» модели рекомендуется проверить статистическую значимость отобранных признаков, а также провести экспертизный анализ признаков. По результатам анализа рекомендуется удалить неинформативные признаки и перестроить «ядро» модели на финальном наборе признаков.

<sup>1</sup> Stone M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions // Journal of the Royal Statistical Society: Series B (Methodological). Vol. 36. No. 2. P. 111-147. DOI: 10.1111/j.2517-6161.1974.tb00994.x.

## LGD-модели для розничного кредитования. Часть 2: разработка «ядра» модели

Рисунок 16

### Метод 5-блочной перекрестной проверки (5-fold cross-validation)



выборка не участвует в данном процессе. Количество блоков для перекрестной проверки определяется аналитиком и зависит от типа данных и специфики задачи.

По умолчанию для настройки гиперпараметров модели используется 5-блочная перекрестная проверка как межотраслевой стандарт в машинном обучении<sup>1</sup>.

### Обучение «ядра» модели

Обучение «ядра» модели осуществляется после подбора гиперпараметров. Процесс обучения «ядра» модели LGD зависит от подхода и алгоритма моделирования. В случае использования подхода Cure/Loss Rate для каждой из оцениваемых компонент строится отдельная модель на отобранных для этой модели наборе признаков. Для моделирования компоненты CR (бинарное событие) рекомендуется использовать логистическую регрессию. Для компоненты LR (непрерывная переменная) используется дерево решений или линейная регрессия.

Проверка качества модели LGD проводится на обучающей и тестовой выборках.

После обучения «ядра» модели рекомендуется проверить статистическую значимость отобранных признаков, а также провести экспертный анализ признаков. По результатам анализа рекомендуется удалить неинформативные признаки и перестроить «ядро» модели на финальном наборе признаков.

<sup>1</sup> Raschka S. (2018). Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning (<https://arxiv.org/abs/1811.12808>).

---

Сергей АФАНАСЬЕВ и др.

---

## Сравнение подходов для разработки «ядра» модели LGD

Сравнение подходов Gross и Cure/Loss Rate проводилось на двух портфелях розничных кредитов, собранных за 6,5 лет:

- 1) кредиты наличными (cash):
  - In-Default: 2 335 805 наблюдений;
  - Will-Default: 481 827 наблюдений;
- 2) кредиты на товары и услуги (POS):
  - In-Default: 1 330 850 наблюдений;
  - Will-Default: 348 857 наблюдений.

Для моделирования компонент использовались классические алгоритмы машинного обучения:

- 1) для подхода Gross:
  - дерево решений (задача регрессии);
  - линейная регрессия;
- 2) для подхода Cure/Loss Rate:
  - логистическая регрессия (для бинарной компоненты CR) + дерево решений (для регрессионных компонент  $LR_c$  и  $LR_t$ );
  - логистическая регрессия (для бинарной компоненты CR) + линейная регрессия (для регрессионных компонент  $LR_c$  и  $LR_t$ ).

Отбор признаков осуществлялся с помощью схемы комбинированного отбора<sup>1</sup>. Настройка гиперпараметров моделей проводилась с помощью 5-блочной перекрестной проверки. Оценка «ядер» моделей делалась на тестовой выборке по двум метрикам качества: Powerstat и MAE (средняя абсолютная ошибка).

### Метрика Powerstat

Для оценки предсказательной способности моделей LGD используется метрика Powerstat, которая рассчитывается как отношение площадей под кривыми фактических и модельных убытков<sup>2</sup>. Расчет метрики Powerstat осуществляется по следующему алгоритму (рис. 17):

1. Для оценки кривой фактических убытков (идеальная модель) наблюдения упорядочиваются по убыванию фактических значений LGD “Will-Default” и LGD “In-Default”. Для каждого наблюдения указывается соответствующая накопленная доля от общего числа наблюдений и от общей суммы фактических значений LGD “Will-Default”

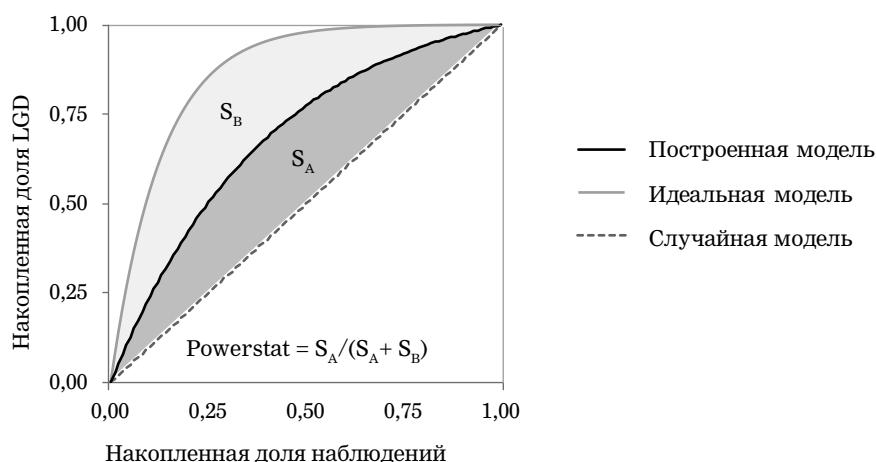
<sup>1</sup> Афанасьев С., Котерева Д., Смирнова А. Two-forest jupt: комбинированный отбор признаков с использованием двухлесосторого метода // Риск-менеджмент в кредитной организации. 2020. № 3.

<sup>2</sup> Li D., Bharik R., Keenan S., Santilli S. (2009). Validation techniques and performance metrics for loss given default models // Journal of Risk Model Validation. Vol. 3. No. 3. P. 3-26. DOI: 10.21314/JRMV.2009.045.

## LGD-модели для розничного кредитования. Часть 2: разработка «ядра» модели

Рисунок 17

### Пример расчета показателя Powerstat через кривые фактических и модельных LGD



и LGD “In-Default”. Далее оценивается площадь ( $S_A + S_B$ ) между построенной кривой фактических убытков и диагональю, соответствующей случайной модели.

2. Для оценки кривой модельных убытков наблюдения упорядочиваются по убыванию модельных значений LGD “Will-Default” и LGD “In-Default”. Для каждого наблюдения указывается соответствующая накопленная доля от общего числа наблюдений и от общей суммы фактических значений LGD “Will-Default” и LGD “In-Default”. Далее оценивается площадь  $S_A$  между построенной кривой модельных убытков и диагональю, соответствующей случайной модели.

3. Показатель Powerstat рассчитывается как отношение площади  $A$  к площади ( $A + B$ ). Чем больше значение показателя Powerstat, тем выше предсказательная способность модели.

Для подхода Cure/Loss Rate метрика Powerstat рассчитывается с учетом всех трех компонент — CR,  $LR_c$  и  $LR_r$ , поскольку компонента CR вносит отдельный вклад в предсказательную способность модели, который не учитывается при индивидуальной оценке компонент  $LR_c$  и  $LR_r$ .

### Результаты экспериментов

В табл. 2 представлены результаты разработанных моделей LGD “In-Default” и LGD “Will-Default” для портфелей Cash и POS.

## Сергей АФАНАСЬЕВ и др.

Таблица 2

### Сравнение подходов и алгоритмов моделирования «ядра» модели LGD (%)

Cash-кредиты		LGD In-Default		LGD Will-Default	
Подход	ML-алгоритм	Powerstat	MAE	Powerstat	MAE
Gross	DecisionTreeRegressor	55,80	10,40	24,38	26,02
	LinearRegression	66,40	10,18	28,57	<b>25,66</b>
Cure/Loss Rate	LogReg + DecisionTreeRegressor	65,91	10,12	27,26	27,39
	LogReg + Linear Regression	<b>66,88</b>	<b>10,05</b>	<b>28,70</b>	27,20
POS-кредиты		LGD In-Default		LGD Will-Default	
Подход	ML-алгоритм	Powerstat	MAE	Powerstat	MAE
Gross	DecisionTreeRegressor	41,68	18,35	14,40	36,81
	LinearRegression	48,22	<b>17,95</b>	<b>20,14</b>	<b>35,27</b>
Cure/Loss Rate	LogReg + DecisionTreeRegressor	47,84	18,04	19,13	35,68
	LogReg + Linear Regression	<b>48,63</b>	17,98	20,04	35,56

**Примечание:** жирным шрифтом выделены результаты лучших моделей.

Подход Cure/Loss Rate показывает лучшие результаты для портфеля cash-кредитов, а также имеет сопоставимое с подходом Gross качество для портфеля POS-кредитов. Также видно, что для моделей “Will-Default” на портфеле POS-кредитов подход Gross на линейной регрессии имеет качество чуть выше, чем подход Cure/Loss Rate. Полученный результат может объясняться тем, что POS-кредиты имеют небольшой срок погашения (в среднем 9 месяцев) и «выздоровление» попавших в дефолт POS-кредитов менее вероятно, чем для более «длинных» cash-кредитов. Кроме того, большая часть дефолтов в POS-кредитах приходится на ранние месяцы графика платежей, в отличие от cash-кредитов, где дефолты распределяются по графику платежей более равномерно.

Дальнейшая разработка модели LGD в рамках цикла CRISP-DM включает в себя:

- калибровку «ядра» модели;
- расчет надбавок консерватизма;
- оценку качества финальной модели.

Эти этапы мы опишем в заключительной части статьи в следующем выпуске журнала. 

В прошлом номере мы рассматривали ежемесячные отчеты<sup>1</sup>, в этом сосредоточимся на отчетах ежеквартальных и немного на макроэкономике. Приведенные в статье примеры отражают состояние реальных кредитных портфелей и позволяют оценить влияние эпидемии COVID-19. В сочетании с прогнозом макроэкономических показателей представленные модели позволяют строить прогноз факторов поведения кредитных портфелей, что необходимо для расчета ожидаемых кредитных убытков в соответствии с МСФО (IFRS) 9 и стресс-тестирования, направленного на конструирование кредитных продуктов и обеспечение устойчивости банка.

## Ежеквартальная отчетность розничных кредитных портфелей: как создать ценность бизнесу? Часть 2

Предлагаемые квартальные отчеты содержат отчеты всех трех уровней (таблица).

### Анализ факторов риска, влияющих на портфель в текущей ситуации (point-in-time)

Основой анализа факторов риска является декомпозиция, детально описанная в журнале ранее<sup>2</sup>, которая позволяет разделить факторы, влияющие на кредиты в зависимости от их возраста, момента выдачи, текущего момента времени и присущих ему характеристик. Все это дополняет результаты анализа усредненных частот переходов кредитов по состояниям, позволяет уточнить их, проанализировать риски отклонения поведения конкретных кредитов от усредненных показателей. Эта декомпозиция соответствует требованиям § 402



Сергей КОПЫЛОВ,  
ООО «Бизнес Системы Консалт», младший партнер, к.ф.-м.н.,  
CFA, FRM

<sup>1</sup> Копылов С. Ежемесячная отчетность розничных кредитных портфелей: как создать ценность бизнесу? Часть 1 // Риск-менеджмент в кредитной организации. 2021. № 3.

<sup>2</sup> Бабиков В. Поведение розничного кредитного портфеля: методики и подходы к оценке // Риск-менеджмент в кредитной организации. 2019. № 4.

## Сергей КОПЫЛОВ

### Таблица

### Квартальные отчеты

Отчет	Назначение	Адресат
Факторы риска, оказывающие влияние на портфель в текущей ситуации (point-in-time)	Информация о неявных особенностях кредитной политики, обуславливающих поведение заемщиков, позволяющая сформировать рекомендации по усовершенствованию скоринга	A, С, В
Макроэкономический анализ	Связь факторов поведения заемщиков с макроэкономической конъюнктурой	A, ALM, МСФО
Планирование портфеля	Объемы портфеля и резервирование в рамках макроэкономических сценариев. При анализе этой отчетности исходя из целевых показателей эффективности кредитования, которые могут быть установлены в декларации аппетита к риску, уточняются планы новых выдач, скоринговые баллы отсечки и пр.	A, ВР, ALM

**Примечание:** для удобства системы отчетов снабжена условным столбцом «Адресат»: А – аналитик кредитных портфелей, ВР – высшее руководство, ALM – управление активами и пассивами, финансовое планирование, ценообразование и балансовые риски, МСФО – подразделение, ответственное за отчетность, В – подразделения подтверждения качества моделей, С – скоринговые карты.

Базеля II<sup>1</sup> и п B5.5.5 МСФО (IFRS) 9, обеспечивая их эффективную реализацию в части управления кредитными портфелями, по которым риск рассчитывается на групповой (в терминах МСФО (IFRS) 9) или портфельной (в терминах Базеля) основе.

Оценим частоты переходов по состояниям, определенным описанным образом, как сочетания действия разных факторов:

$$R(t_1, t_2) = R_0(t_1 - t_2) + \mu_1(t_1)M + \xi(t_2)Q + \mu_2(t_1)MM + \pi(t_1)PP,$$

где  $\mu_1(t_1)$  — коэффициент, характеризующий силу влияния фактора макроэкономики, который прогнозируется в будущем;

$\mu_2(t_1)$  — коэффициент, характеризующий силу влияния фактора сбора просроченной задолженности, который прогнозируется в будущем;

<sup>1</sup> bcbs128: International Convergence of Capital Measurement and Capital Standards. A Revised Framework // Basel Committee on Banking Supervision. Basel, 2006.

## Ежеквартальная отчетность розничных кредитных портфелей: как создать ценность бизнесу? Часть 2

$\xi(t_2)$  — коэффициент, характеризующий качество поколения кредитов, выданных в момент времени  $t_2$ ;

$\pi(t_1)$  — коэффициент, характеризующий силу влияния фактора досрочного погашения, который прогнозируется в будущем;

$t_0$  — месяц, по результатам которого рассчитываются ожидаемые кредитные убытки;

$t_1$  — месяц, за который рассчитываются различные характеристики групп кредитов;

$t_2$  — месяц, в который выдано определенное поколение кредитов;

$M$  — матрица фактора макроэкономики, все элементы которой в стандартной спецификации равны нулю, кроме элемента, равного 1, соответствующего переходу из состояния отсутствия просрочки (глубина просрочки равна 0) в состояние просрочки 1 месяц;

$PP$  — матрица фактора досрочного погашения, все элементы которой в стандартной спецификации равны нулю, кроме элемента, соответствующего переходу из состояния отсутствия просрочки в состояние выплаты по кредитам (обозначаемое индексом 100);

$Q$  — матрица фактора качества поколения кредитов, все элементы которой в стандартной спецификации равны нулю, кроме элемента, равного 1, соответствующего переходу из состояния отсутствия просрочки (глубина просрочки равна 0) в состояние просрочки 1 месяц;

$MM$  — матрица фактора сбора просроченной задолженности, все элементы которой равны нулю, кроме элементов, соответствующих переходам, увеличивающим глубину просрочки на 1 месяц.

Данная декомпозиция позволяет разделить влияние различных макроэкономических и портфельных факторов.

Результаты приведенных на рис. 18 и 19 графиков имеют наглядную интерпретацию, которая позволяет как наметить пути совершенствования кредитного процесса, так и обеспечить более качественное планирование ресурсов банка с учетом макроэкономического прогнозирования и оценки макроэкономических рисков. Графики приведенного примера (данные которого, хотя они и модельные, отражают состояние реальных кредитных портфелей) позволяют оценить влияние пандемии COVID-19:

— весной – в июне 2020 г. резко выросла добавочная (к естественной) частота просрочки по кредитам;

— одновременно выросли частоты углубления просрочки: если заемщики в нее попадали, возвращаться к обслуживанию кредитов им было сложнее;

— падение объемов добавочного досрочного погашения кредитов не выбивалось из обычных сезонных трендов;

## Сергей КОПЫЛОВ

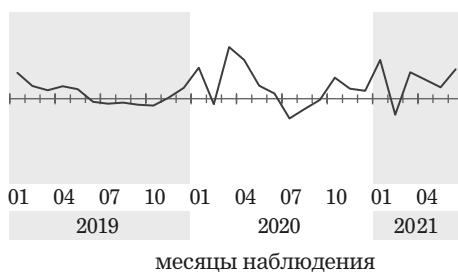
Рисунок 18

### Факторы декомпозиции по кредитному портфелю: пример (фактор Q – по месяцам выдач, основные факторы – по месяцам наблюдения)

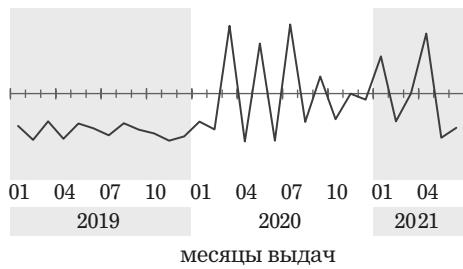
Макроэкономика (фактор M):  
дополнительная частота переходов  $0 \rightarrow 1$



Макроэкономика и сбор просрочки (фактор MM):  
доп. частота углубления просрочки



Качество новых выдач (фактор Q):  
дополнительная частота переходов  $1 \rightarrow 2$



Досрочное погашение (фактор P):  
дополнительная частота переходов  $0 \rightarrow 100$

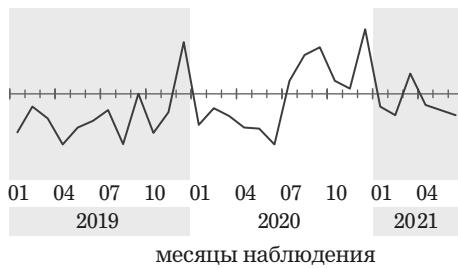


Рисунок 19

### Фактор декомпозиции по кредитному портфелю: пример

Качество начальных выдач (фактор QP):  
% дополнительная частота переходов  $0 \rightarrow 100$



## Ежеквартальная отчетность розничных кредитных портфелей: как создать ценность бизнесу? Часть 2

— с июля 2020 г., когда были свернуты меры по изоляции населения, тут же упали (хотя и остались положительными) дополнительные частоты углубления просрочки, но резко усилилось досрочное погашение: заемщики стали усиленно делать сбережения из предосторожности и освобождаться от кредитов;

— особое внимание банкам следует уделять переработке моделей принятия кредитного риска (скоринговых карт): высокая волатильность влияния качества новых выдач на добавочную частоту углубления просрочки (фактора  $Q$ ) является тревожным сигналом и наблюдается во многих реальных кредитных портфелях.

Анализ фактора  $MM$  в сочетании с информацией о затратах и методах взыскания просроченной задолженности позволяет измерить эффективность этого процесса.

### Макроэкономический анализ

Макроэкономические зависимости представляются в следующем виде:

$$M_t = 2,03 - 0,052 \times \log_{1,1} Brent_t + \varepsilon_t,$$

$$P_t = 1,43 - 0,353 \times MosPrime_t + 0,402 \times RT_t + 1,412 \times S_{12} + \varepsilon_t,$$

$$MM_t = 25,23 - 0,602 \times Brent_t + 38,24 \times BSK_t + 3,982 \times S_{12} + \varepsilon_t.$$

Под коэффициентами указаны соответствующие  $t$ -статистики, звездочкой (\*) помечено значение  $t$ -статистики, соответствующее статистической значимости на доверительном уровне 0,9. Без звездочки показаны значения, соответствующие более высоким доверительным уровням, нежели 0,95 (этот доверительный уровень обычно используется в исследованиях подобного рода). Коэффициенты, описывающие зависимость поведения портфеля от макроэкономических параметров, приведены исключительно для наглядности и в других портфелях могут отличаться так же, как и перечень макроэкономических факторов.

В данном примере:

$\log_{1,1} Brent_t$  — логарифм по основанию 1,1 среднемесячной цены на нефть марки Brent, выраженной в долларах США за 1 нефтяной баррель;

$MosPrime_t$  — среднемесячная процентная ставка MosPrime 3M;

$RT_t$  — индекс изменения оборота розничной торговли в Российской Федерации, рассчитанный как логарифм по основанию 1,1 отношения оборота розничной торговли в месяц  $t$  к обороту розничной торговли в месяц  $t - 1$ , скорректированный на сезонность;

## Сергей КОПЫЛОВ

$BSK_t$  — индекс изменения стоимости определенной Росстатом товарной корзины, представляющий собой логарифм по основанию 1,1 отношения цен этой корзины в месяц  $t$  и в месяц  $t - 1$ ;

$S_1$  — величина, принимающая значение 1 в январе и 0 в иные месяцы;

$S_{12}$  — величина, принимающая значение 1 в декабре и 0 в иные месяцы.

Тем не менее, поведение других кредитных портфелей может зависеть от других макроэкономических факторов. Например, изредка встречается майская сезонность.

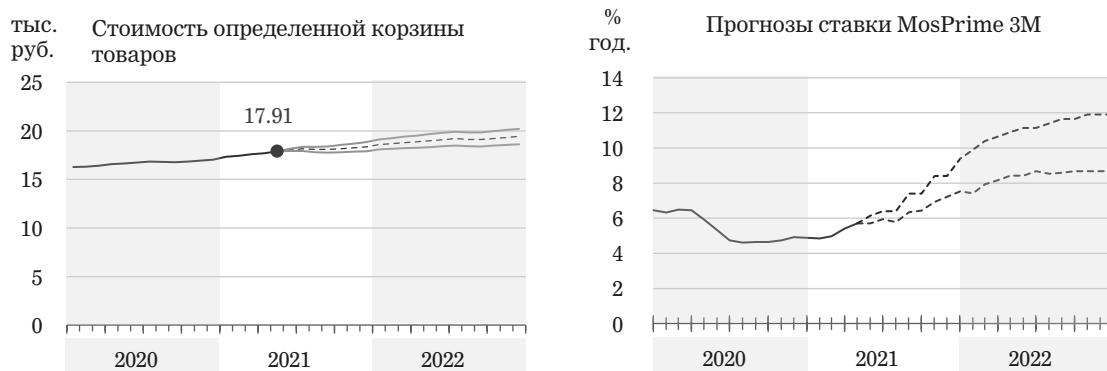
Очень часто поведению ошибок (обозначены  $\varepsilon_t$ ) свойственна авторегрессия: они подчиняются процессам ARMA (2, 2) (большие порядки в практике не встречаются).

В сочетании с прогнозом макроэкономических показателей представленные модели позволяют строить прогноз факторов поведения кредитных портфелей.

В рамках отчета необходимо строить два варианта прогноза: ожидаемое и стрессовое изменение факторов поведения портфелей. Стressовое поведение факторов портфеля обычно соответствует худшей с точки зрения поведения портфеля квантили значения макроэкономической переменной. Целевая вероятность квантили принимается равной 0,95. Примеры макроэкономических прогнозов приведены на рис. 20, примеры прогнозов факторов портфеля — на рис. 21.

Рисунок 20

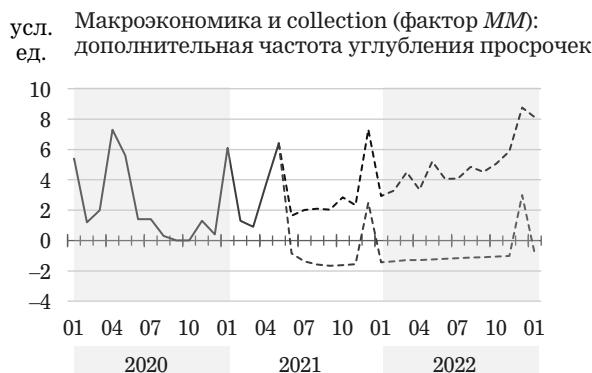
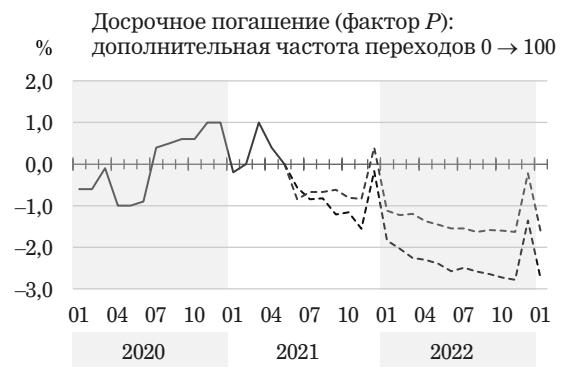
### Прогнозирование стоимости потребительской корзины и ставки MosPrime 3M



## Ежеквартальная отчетность розничных кредитных портфелей: как создать ценность бизнесу? Часть 2

Рисунок 21

### Прогнозы факторов портфеля



### Планирование портфеля

Для планирования портфеля наряду с прогнозом характеристик риска необходима информация о планируемых объемах новых выдач. Математические методы планирования должны решать две задачи:

- 1) как от целевой структуры портфеля перейти к плановому заданию;
- 2) как от новых сделок, их объемов и характеристик риска перейти к описанию портфеля, спрогнозировать его структуру.

Методология динамического баланса позволяет оперативно рассчитывать последствия управленических решений (рис. 22), формируя отчеты уровня «прогноз».

Отчетность о планировании портфеля начинается с планов по новым выдачам (рис. 23).

## Сергей КОПЫЛОВ

Рисунок 22

### Методология динамического баланса



Рисунок 23

### Сценарии новых выдач



## Ежеквартальная отчетность розничных кредитных портфелей: как создать ценность бизнесу? Часть 2

Рисунок 24

### Распределение портфеля по срокам просрочки (а) без включения и (б) с включением в портфель кредитов из риск-класса 13

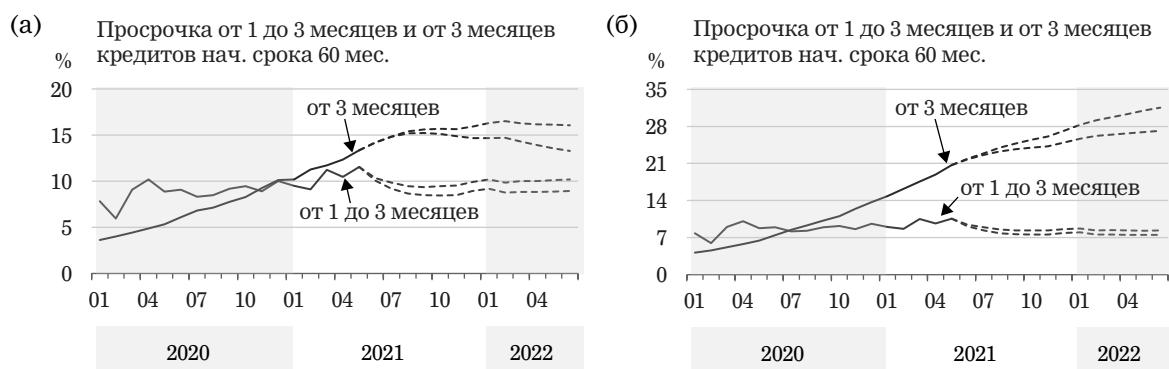
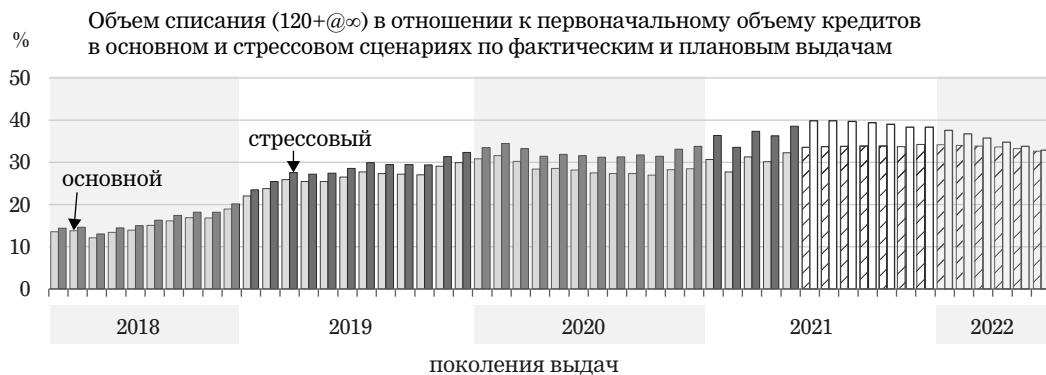


Рисунок 25

### Винтажные характеристики портфеля



Сценарии макроэкономических показателей обеспечивают построение совместной отчетности по объемам просроченной задолженности в кредитном портфеле (рис. 24), винтажным характеристикам (рис. 25), резервам (рис. 26, 27). Совместный анализ исторических и прогнозных данных, результатов стресс-тестирования позволяет сформировать исчерпывающую картину и принять управленические решения.

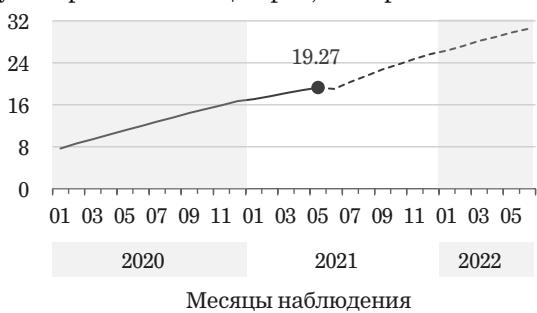
Анализ представленной отчетности фактически инициирует процессы изменения планов кредитования (рис. 28 и 29). Скорректированные результаты метрик риска портфеля приведены на рис. 30 и 31,

## Сергей КОПЫЛОВ

Рисунок 26

### Резервы в соответствии с МСФО (IFRS) 9 в рамках (а) основного и (б) стрессового сценариев

(а) млрд руб. Резервы в соответствии с МСФО (IFRS) 9 при основном сценарии, нач. срок 60 мес.



(б) млрд руб. Резервы в соответствии с МСФО (IFRS) 9 при стрессовом сценарии, нач. срок 60 мес.

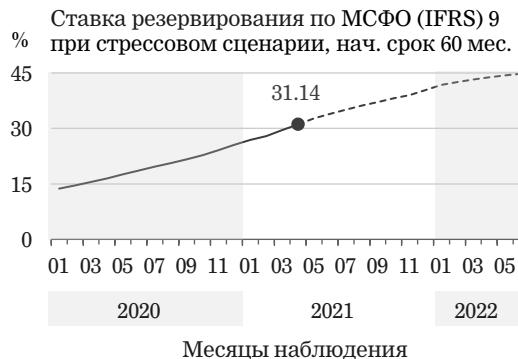
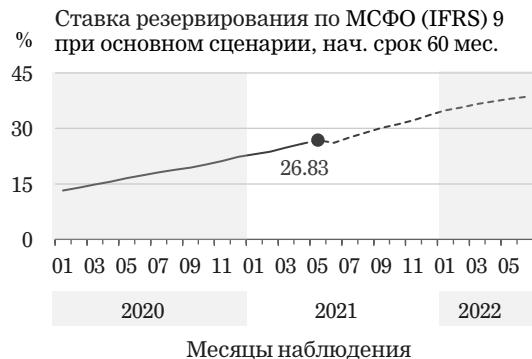
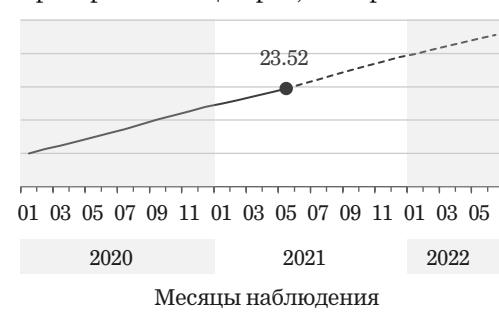
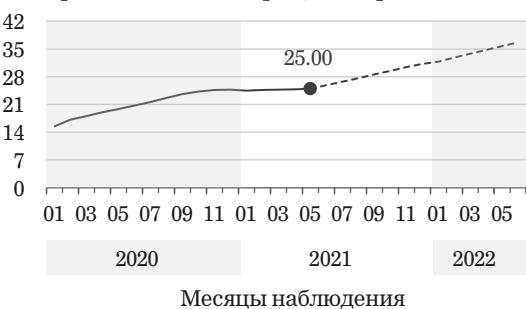


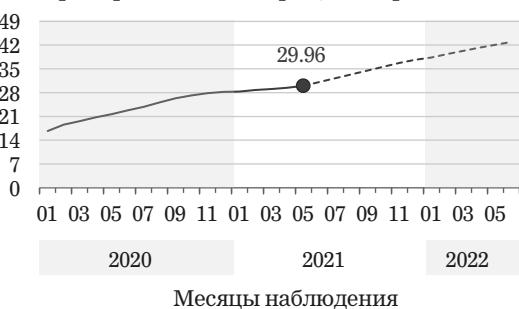
Рисунок 27

### Резервы в соответствии с МСФО (IFRS) 9 при расчете на весь срок кредитов в рамках (а) основного и (б) стрессового сценариев

(а) млрд руб. Резервы на весь срок кредитов при основном сценарии, нач. срок 60 мес.



(б) млрд руб. Резервы на весь срок кредитов при стрессовом сценарии, нач. срок 60 мес.



управление кредитным портфелем \ МСФО (IFRS) 9 \ ожидаемые кредитные убытки

## Ежеквартальная отчетность розничных кредитных портфелей: как создать ценность бизнесу? Часть 2

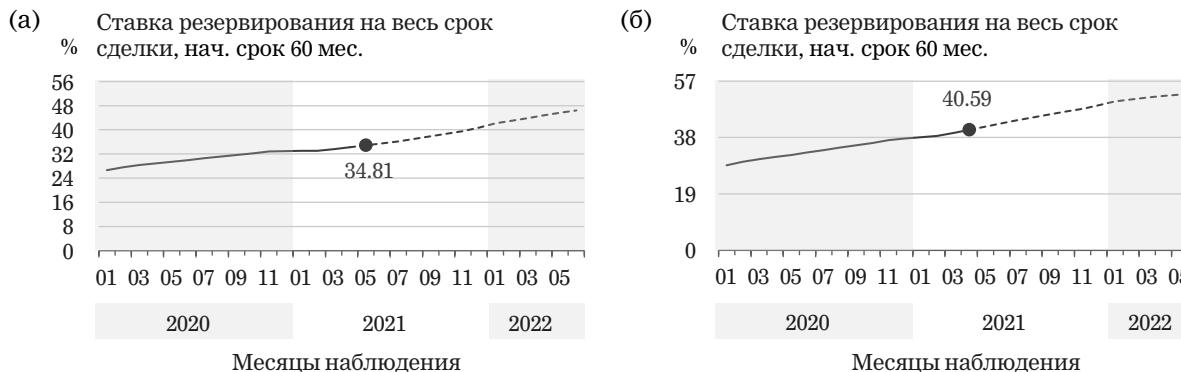


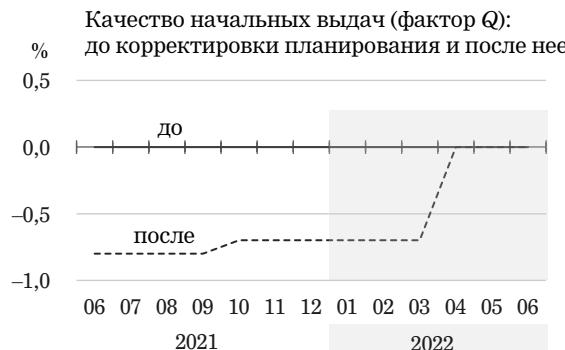
Рисунок 28

### Объемы новых выдач, скорректированные в рамках процедур планирования, инициированных в результате анализа отчетов



Рисунок 29

### Начальное кредитное качество портфеля, скорректированное в рамках планирования



## Сергей КОПЫЛОВ

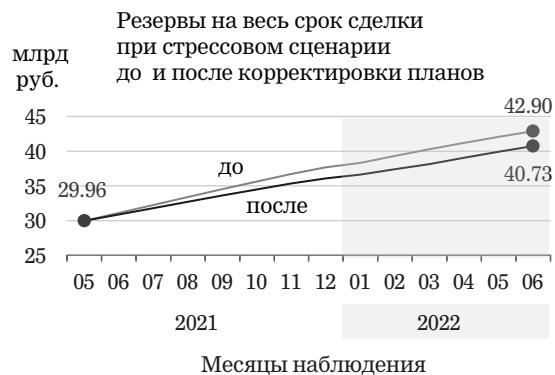
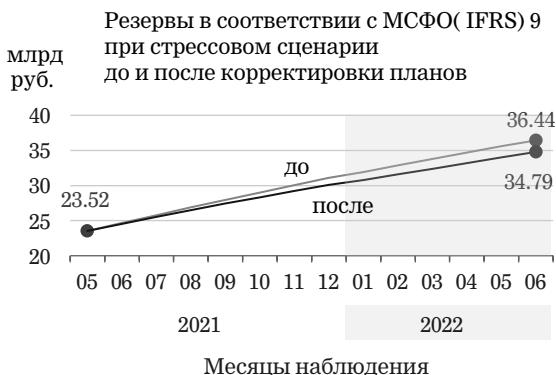
Рисунок 30

### Винтажные характеристики планируемых выдач до и после пересмотра плана



Рисунок 31

### Резервы в соответствии с МСФО (IFRS) 9 в рамках стрессового сценария



аналогичных рис. 24–27 соответственно. Все это демонстрирует, как развернутая аналитическая отчетность позволяет повысить эффективность планирования.

При этом следует учесть, что планирование — это организационная процедура, реализуемая посредством формализованного диалога между бизнес-подразделениями, финансовым департаментом, внутренним казначейством, риск-менеджментом. Суть этого диалога — сформировать достижимые планы операций, приемлемые с точки зрения целевых значений нормативов и иных ограничений риска.

---

В банковской системе есть сервисы ниже «ватерлинии», которыми почему-то не принято пользоваться. В этой статье мы подключим сервис интеграции «ЗаЧестныБизнес» (№ 10 в рейтинге «Эксперт РА» и лучший балл по стоимости<sup>1</sup>) по цене ежемесячного утреннего кофе для главного редактора и посмотрим: может, все-таки дать шанс бюджетным решениям и сэкономить миллионы рублей в месяц кредитной организации, которая привыкла считать деньги?

## От СПАРК до Creditnet: глубинный обзор систем анализа контрагентов. «ЗаЧестныБизнес»

«ЗаЧестныБизнес» предлагает два варианта использования сервиса: через олдскульный php-интерфейс, насыщенный рекламой и всплывающими окнами, и API. Цена — от 5970 руб. в месяц за 3000 запросов (запрос тарифицируется по применяемому методу, а не компании). Мы, помня, что исходник всех сервисов проверки контрагентов один и лежит где-то в ЦОД налоговой службы, выбрали второй вариант, чтобы проверить: можно ли сильно сэкономить на данных без потери качества?

Честно отложенные деньги от утреннего кофе, кажется, уже окунулись: сервис имеет удобный личный кабинет и автоматически выдает API-ключи и отчетные документы. Итак, мы приступили к тестированию уже через несколько минут после оплаты. Никаких длительных переписок, выяснения обстоятельств, прайсинга, соглашений. То, что нужно!

Основное назначение тестов — создание датасета для корпоративной модели. Для этого из реестра залогов выгрузим 500 клиентов Сбера и на них протестируем сервис.

### Методы проверки

Сервис предлагает некие базовые методы. Главный из них — «Основные сведения (card)» — соответствует GetCompanyExtendedReport<sup>2</sup>

---

<sup>1</sup> От СПАРК до Creditnet: глубинный обзор систем анализа контрагентов. «СКРИН» // Риск-менеджмент в кредитной организации. 2021. № 1.

<sup>2</sup> Документация: <http://sparkgatetest.interfax.ru/iFaxWebService/>.

---

у СПАРК и возвращает довольно большой набор полей, включающий практически все, к чему мы привыкли в сервисах проверки контрагентов.

Например, по средней строительной компании «ЭнергоСила» нормализованный JSON вернул 185 полей, многие из которых вложенные. В отличие от СПАРК, можно выбрать JSON или XML, установив параметр `_format — JSON`. Это крайне полезно для ускоренной разработки модели в Python-среде, в которой Pandas довольно успешно распознает json-схемы из строкового значения.

Из-за вложенности приходится извлекать поля методами типа:

```
url_arbitr = "https://zachestnyibiznesapi.ru/paid/data/court-
arbitration?id="+inn+api_key
session = requests.Session()
headers = {"accept": "application/json", "content-type": "application/json"}
response = session.get(url_arbitr, headers=headers, verify=False)
data_arbitr = json.loads(response.text)
df_arbitr = pd.json_normalize(data_arbitr['body'])
inn = pd.DataFrame(pd.DataFrame(df_arbitr['docs'][0][0]['точно']
['дела']).T['Истец'][0])['ИНН'][0]
```

Но в целом к дизайну методов можно привыкнуть. Конечно, в отличие от СПАРК, который предлагает разнообразные точечные методы проверки и получения информации из специфических источников, сервис содержит ограниченное количество методов. Но для построения корпоративной модели в целом они подойдут — нам, как и в большинстве сервисов, отдадут данные о финансовой отчетности, арбитражах, исполнительных листах, госзакупках и многое нефинансовой информации из ЕГРЮЛ, которая в основном не несет полезной нагрузки, если только вы не занимаетесь графиками.

Ключевые методы:

---

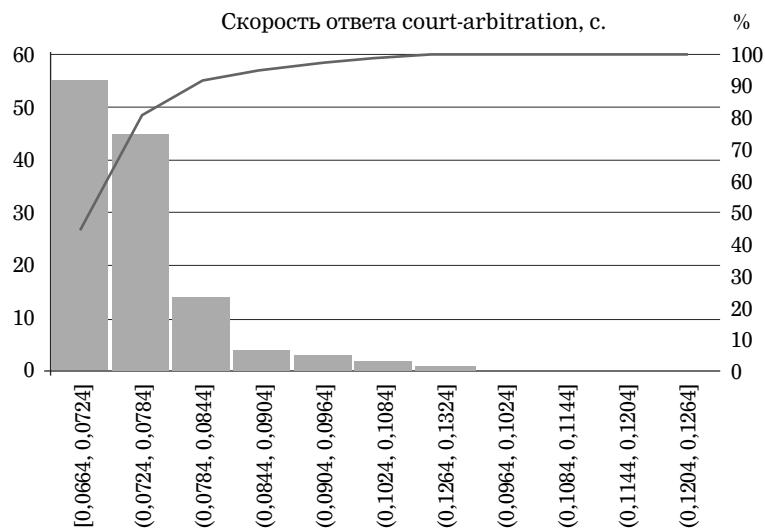
Основные сведения (card)
Фин. отчетность (fs)
Фин. отчетность ФНС (fs-fns)
ФССП (fssp-list)
Судебные дела (court-arbitration)
Контакты (contacts)
Проверки (proverki)
ТОП закупок (zakupki-top)
Список сообщений по банкротству (bankrupt-message-list)
Список лизинговых договоров (lease-contracts-list)
Связанные организации (affilation-company)

---

## От СПАРК до Creditnet: глубинный обзор систем анализа контрагентов. «ЗачестныйБизнес»

Чтобы избежать проблему нулей в качестве первого символа ИНН для ряда регионов, в сервисе в основном используется ОГРН как идентификатор компании.

Время ответа в целом стабильно и составляет менее 1 секунды по запросу (основная часть запросов проходит менее чем за 0,1 с.).



В разумного размера батчах блокировок не выявлено, что радует.

Метод «Основные сведения» удачно продуман с точки зрения экономии вызовов и возвращает большинство исторических данных в одном запросе, что удобно, если вы привыкли генерировать большое количество признаков разом и далее выбирать из них релевантные.

Пример с отчетностью (в данном методе она доступна на верхнем уровне):

```
Status: 200 OK    Time: 141 ms
PRETTY  SHOW HEADERS  DARK THEME 

80      "mass_owners": "0",
81      "mass_leaders": "0"
82  },
83  "Ф02020": {
84    "выручка": 1669272000,
85    "прибыль": 88592000,
86    "оснсредства": 99826000,
87    "кредиторзадолжн": 925541000,
88    "дебиторзадолжн": 626250000
89  },
90  "Ф02019": {
91    "выручка": 1644552000,
92    "прибыль": 82840000,
93    "оснсредства": 131972000,
94    "кредиторзадолжн": 1018263000,
95    "дебиторзадолжн": 742327000
96  },
97  "Ф02018": {
98    "выручка": 1372429000,
99    "прибыль": 62062000,
100


```

Любопытная деталь: сервис восстанавливает уровень ФОТ и среднюю заработную плату, очевидно, из страховых взносов. Проведем расчеты сами, исходя из того, что ставка по взносам в Фонд обязательного медицинского страхования составляет 5,1% от ФОТ, а взносы ФНС любезно раскрывает всем желающим:

Год	Взносы в Фонд обязательного медицинского страхования, руб.	Число сотрудников	ФОТ, руб.	Средняя заработная плата, руб.
2020	7 627 976,00	150	149 568 157	83 093
2019	6 664 325,00	150	130 673 039	72 596
2018	4 900 228,00	222	96 082 902	36 067
2017	2 899 201,00	175	56 847 078	27 070

С сервисом все сошлось.

Дополнительные особенности методов:

- 1) при отсутствии данных структура ответа меняется (поле body отсутствует);
- 2) запрос по судебным делам требует не ОГРН, а ИНН;
- 3) метод поиска исполнительных листов не работает для индивидуальных предпринимателей;
- 4) в среде Python некоторые методы выдают невалидный JSON из-за escape-символов (“Наименование”: “МУНИЦИПАЛЬНОЕ КАЗЕННОЕ ПРЕДПРИЯТИЕ \\\" СПЕЦАВТОДОР\\””), что создает проблемы при сериализации;
- 5) в суммах арбитражных споров округление до целого, в отличие от, например, API «Контур.Фокуса»;
- 6) в возвращаемых ответах имеются загадочные поля вроде ‘Статус’: 1 (по всей вероятности, историческая информация, не актуальная на сегодня).

Перейдем к главному.

## Консистентность данных

Консистентность данных финансовой отчетности мы проверили на 500 компаниях из портфеля Сбера.

Структура данных довольно неудобна. Такое ощущение, что API написан «для себя» и не для целей ML, ведь баланс — это по определению таблица, а здесь встречается до 5–6 вложенных сущностей и активно используются атрибуты. Дополнительные особенности:

- 1) реализованы два разных метода для запроса балансов до 2018 г. и после;
- 2) балансы 2019 и 2020 гг. не содержат кодов строк;
- 3) балансы МСП отличаются по структуре от обычных балансов;

## От СПАРК до Creditnet: глубинный обзор систем анализа контрагентов. «ЗаЧестныЙБизнес»

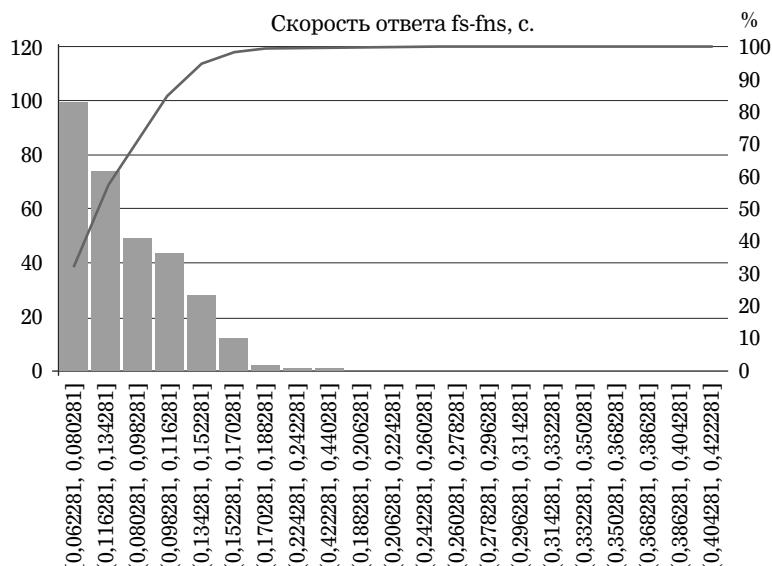
- 4) сама структура не содержит кодов балансов;
- 5) структура крайне запутана. Например, мы пытаемся получить строку «Денежные средства» в случае баланса МСП:

```
pd.json_normalize(data_fns['body']['Документ']['Баланс']['Актив']['ДенежнСр'])
```

А вот так в случае обычного баланса:

```
pd.json_normalize(data_fns['body']['Документ']['Баланс']['Актив']['ОбА']
['ДенежнСр'])
```

В сервисе почему-то отсутствует отчетность по ИНН 7707840279, хотя на официальном ресурсе ФНС данная компания раскрыта<sup>1</sup>. Но в целом отсутствие расхождений по всему объему проверяемых компаний радует. Скорость ответа нестабильна, и на нашем мини-портфеле в 500 компаний вкупе с сериализацией сервис немного «тормозит»:



Арбитражная статистика:

```
Status: 200 OK Time: 141 ms
PRETTY SHOW HEADERS DARK THEME ^

633      "СуммоНедоимЗадолжст": [],
634      "НедобросовПостав": 0,
635      "СудыСтатистика": {
636        "истец": {
637          "Сумма": 211506297,
638          "Количество": 64
639        },
640        "ответчик": {
641          "Сумма": 81364005,
642          "Количество": 38
643        }
644      },
```

<sup>1</sup> <https://bo.nalog.ru/organizations-card/8710687>.

не сходится с данными kad.arbitr.ru:

ЭЛЕКТРОННОЕ ПРАВОСУДИЕ	ЭЛЕКТРОННОЕ ПРАВОСУДИЕ
Фильтр дела Участник дела Судья Суд Номер дела Дата регистрации дела Найти Найдено 61 дел	Участник дела Ответчик Судья Суд Номер дела Дата регистрации дела Найти Найдено 35 дел

Данные о госзакупках:

```
Status: 200 OK   Time: 141 ms
PRETTY  SHOW HEADERS  DARK THEME 
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702 } }
```

также не сходятся:

Реестр контрактов, заключенных заказчиками																							
Введите полностью или часть номера реестровой записи контракта, идентификационного кода закупки (ИКЗ), наименования Заказчика, номера контракта, предмет... <input type="button" value="Искать точно, как в запросе"/>																							
<b>Результаты поиска</b> 60 записей <table border="1"> <thead> <tr> <th colspan="2">Параметры поиска</th> </tr> </thead> <tbody> <tr> <td><input type="checkbox"/> Открыть дополнительную информацию</td> <td>Сортировать по: <input type="button" value="Дата обновления"/></td> </tr> <tr> <td colspan="2"> <input type="checkbox"/> Все параметры →         </td> </tr> <tr> <td colspan="2"> <b>Закон</b> <input checked="" type="checkbox"/> 44-ФЗ           <input checked="" type="checkbox"/> 94-ОЗ         </td> </tr> <tr> <td colspan="2"> <b>Статус контракта</b> <input checked="" type="checkbox"/> Исполнение           <input checked="" type="checkbox"/> Исполнение завершено           <input checked="" type="checkbox"/> Исполнение прекращено         </td> </tr> <tr> <td colspan="2"> <b>Заказчик</b>           АДМИНИСТРАЦИЯ МУНИЦИПАЛЬНОГО ОБРАЗОВАНИЯ ТОРГОВЫЙ ОКРУГ "НОГЛИКСКИЙ"         </td> </tr> <tr> <td colspan="2"> <b>Контракт</b>           № АС-01         </td> </tr> <tr> <td colspan="2"> <b>Объекты закупки</b>           Реконструкция системы водотведения пгт. Ноглики         </td> </tr> <tr> <td colspan="2"> <b>Срок исполнения</b>           30.12.2022         </td> </tr> <tr> <td colspan="2"> <b>Размещение контракта в реестре контрактов</b>           26.09.2018         </td> </tr> <tr> <td colspan="2"> <b>Обновлен контракт в реестре контрактов</b>           25.11.2021         </td> </tr> </tbody> </table>		Параметры поиска		<input type="checkbox"/> Открыть дополнительную информацию	Сортировать по: <input type="button" value="Дата обновления"/>	<input type="checkbox"/> Все параметры →		<b>Закон</b> <input checked="" type="checkbox"/> 44-ФЗ <input checked="" type="checkbox"/> 94-ОЗ		<b>Статус контракта</b> <input checked="" type="checkbox"/> Исполнение <input checked="" type="checkbox"/> Исполнение завершено <input checked="" type="checkbox"/> Исполнение прекращено		<b>Заказчик</b> АДМИНИСТРАЦИЯ МУНИЦИПАЛЬНОГО ОБРАЗОВАНИЯ ТОРГОВЫЙ ОКРУГ "НОГЛИКСКИЙ"		<b>Контракт</b> № АС-01		<b>Объекты закупки</b> Реконструкция системы водотведения пгт. Ноглики		<b>Срок исполнения</b> 30.12.2022		<b>Размещение контракта в реестре контрактов</b> 26.09.2018		<b>Обновлен контракт в реестре контрактов</b> 25.11.2021	
Параметры поиска																							
<input type="checkbox"/> Открыть дополнительную информацию	Сортировать по: <input type="button" value="Дата обновления"/>																						
<input type="checkbox"/> Все параметры →																							
<b>Закон</b> <input checked="" type="checkbox"/> 44-ФЗ <input checked="" type="checkbox"/> 94-ОЗ																							
<b>Статус контракта</b> <input checked="" type="checkbox"/> Исполнение <input checked="" type="checkbox"/> Исполнение завершено <input checked="" type="checkbox"/> Исполнение прекращено																							
<b>Заказчик</b> АДМИНИСТРАЦИЯ МУНИЦИПАЛЬНОГО ОБРАЗОВАНИЯ ТОРГОВЫЙ ОКРУГ "НОГЛИКСКИЙ"																							
<b>Контракт</b> № АС-01																							
<b>Объекты закупки</b> Реконструкция системы водотведения пгт. Ноглики																							
<b>Срок исполнения</b> 30.12.2022																							
<b>Размещение контракта в реестре контрактов</b> 26.09.2018																							
<b>Обновлен контракт в реестре контрактов</b> 25.11.2021																							

## От СПАРК до Creditnet: глубинный обзор систем анализа контрагентов. «ЗачестныйБизнес»

Реестр договоров, заключенных заказчиками по результатам закупки

Введите наименование (номер) договора, наименование предмета договора, наименование или ИНН заказчика

Искать точно, как в запросе

Результаты поиска 7 записей

Сортировать по: <input type="button" value=""/> Дата обновления <input type="button" value=""/>		Параметры поиска	
Все параметры <input type="button" value=""/>			
Статус договора			
<input checked="" type="checkbox"/> Исполнение			
<input checked="" type="checkbox"/> Исполнение завершено			
<input checked="" type="checkbox"/> Исполнение прекращено - расторжение			
<input checked="" type="checkbox"/> Исполнение прекращено - окончание срока действия договора			
Nº 56500000024160004030000	Исполнение	Цена договора 7 500 000,00 ₽	
Номер договора 1106-18/16		Заключение договора 27.12.2016	
Заказчик ПУБЛИЧНОЕ АКЦИОНЕРНОЕ ОБЩЕСТВО ЭНЕРГЕТИКИ И ЭЛЕКТРИФИКАЦИИ "САХАЛИНЭНЕРГО"		Срок исполнения 25.11.2016 – 30.12.2016	
		Размещено 30.12.2016	Обновлено 26.02.2019

Однако с такими данными в сервисах это часто случается.

Поисковик по текстовым судебным документам непростой. Многие игроки, вероятно, не используют официальный API от pravo.ru, ftp сайта госзакупок также довольно запутанный. Поэтому полного совпадения информации в разных сервисах нет: если совпадение хотя бы в пределах 10% — это уже успех. Аналитики привыкли, что «Контур.Фокус» «не бьется» со СПАРК, тут же представлен третий вариант «истины».

Мы решили немного отвлечься от портфеля Сбера и посмотреть, что сервис выдаст на тему исполнительных листов крупной публичной компании.

Известно, что в данных ФССП много «мусора» (несоответствие наименований, некорректное использование кавычек в названии, отсутствие идентификаторов), с которым, впрочем, большинство сервисов научились более-менее корректно работать.

В ответ на запрос по НК «Роснефть» мы получили совсем удивительные результаты: три исторических штрафа ГИБДД (очевидно, что исторические данные в сервисе как-то чистятся, но тогда зачем их выдавать? ведь на самом деле их сотни) и один текущий штраф в Чечне.

Вопрос о соответствии данным fssp.gov.ru уже не стоит — данные, разумеется, не соответствуют (пропущены два неоплаченных штрафа на сумму около 0,5 млн руб., а указанный «чеченский» штраф в системе отсутствует).

С крупной (по меркам Сахалина) строительной компанией случился иной фейл. Сервис выдал, что она на УСН:

Status: 200 OK Time: 141 ms PRETTY SHOW HEADERS DARK THEME

```
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
1,
  "НалогРежим": "Упрощенная система налогообложения (УСН)",
  "ОсновПоказОтчетнис": [
    {
      "Год": 2017,
      "СумДоход": "973908000.00",
      "СумРасход": "918374000.00"
    },
    {
      "Год": 2019,
      "СумДоход": "1658927000.00",
      "СумРасход": "1552392000.00"
    }
  ],
  "ОсновПоказОтчетн": [
    {
      "СумДоход": "1658927000.00",
      "СумРасход": "1552392000.00"
    }
  ],
  "УплачСтрахВзносист": [
```

Официальные данные налоговой службы другие:

Специальные налоговые режимы: Специальный налоговый режим не применяется

данные на 31 декабря 2020

2020	2019	2018
Наименование специального налогового режима		
Специальный налоговый режим не применяется		

И это уже критичная проблема.

Скупой платит дважды. Сэкономленные деньги пришлось потратить на кофе двойной крепости, жизненно необходимый для сотрудника, который проверял многочисленные ошибки платного сервиса. 

---

Компания QIWI активно развивает бизнес, основанный на аналитике больших данных. Опыт QIWI говорит о том, что оттачивать лишь одну модель не имеет смысла, важно создать конвейерный подход. Расскажем о топ-6 факторов, влияющих на модельный риск, и трех китах прозрачного процесса разработки. А также о том, как при помощи минимального набора методов построить модели с life time более двух лет.

## Лучшие практики data science проектов для управления модельным риском: опыт QIWI

Как показала наша практика, большинство факторов, влияющих на модельный риск (табл. 1), связаны с процессом производства и применения моделей.

На наш взгляд, прозрачность разработки модели на каждой стадии важнее числа тестировщиков и объема документации.

### Три кита разработки

С точки зрения датасаентиста эволюция модели выглядит так:

- 1) появляется работающий прототип в блокноте;
- 2) мы автоматизируем модель;
- 3) начинаются проблемы в проде.

Что с этим делать?

Чтобы неожиданностей в процессе эксплуатации модели не случалось, важно уделять внимание трем этапам создания моделей (табл. 2).

Основной вопрос на этих стадиях — наличие платформы в компании (у нас реализована платформа разработки, деплоя и тестирования). Такие модные слова, как Kubernetes, микросервисы и пр., хороши только в случае, если у вас достаточно людей, чтобы все это обслуживать. Если да, то нужно понимать уже на старте, кто и в какой роли будет заниматься обслуживанием моделей и процессов поставки данных. Важно отметить, что штат непосредственно на разработку



Лидия ХРАМОВА,  
QIWI, *team lead data  
scientist*

## Лидия ХРАМОВА

Таблица 1

### Топ-6 факторов, влияющих на модельный риск

Воспроизводимость	Предсказания модели всегда можно воспроизвести. Если не получается повторить результат модели — это первое, с чем стоит поработать
Сегмент/аудитория	Модель обучается и применяется на одних и тех же сегментах
Таргет	Модель применяется для оценки релевантного целевого события
Входящие данные	Контролируются полнота, распределение и содержание входящих данных. Это не только внешние данные, но и ваши собственные внутренние данные. Нужно отслеживать их качество не только простыми метриками (есть данные или нет), но и более сложными, которые, например, показывают, что изменилась структура распределения данных. Garbage in — garbage out: первое правило моделей
Жизненный цикл	Настроен прозрачный процесс разработки, изменения и вывода из эксплуатации данных и моделей. Много документации не нужно. Но важно настроить прозрачный процесс: кто и как разрабатывает модели, как проходит приемка, кто их тестирует, как они выводятся в производство и самое важное — как они выходят из эксплуатации, потому что у каждой модели есть срок жизни и однажды наступает пора ее менять
Правила применения	Настроен прозрачный процесс расчета и изменения cut off

Таблица 2

### Советы для старта

Проверка гипотезы (MVP)	Deploy	Эксплуатация
<ul style="list-style-type: none"> <li>Определите аудиторию и таргет.</li> <li>Начните с baseline: 80% успеха на этапе MVP — глубокое понимание данных и образа результата. Сложные модели дают не такой большой прирост, начинайте с простого.</li> <li>Лучше оставить 50–70 фич вместо 5000 разных категорий, чтобы увидеть больше аспектов жизни заемщика. Признаки мы отбираем сами в зависимости от продукта и качества признака по Gini, при этом пользуемся только внутренними данными.</li> <li>Для каждого сегмента надо найти свой актуальный набор признаков.</li> <li>Сохраните все артефакты: модель, обучающую выборку, предсказания модели, собранные признаки.</li> <li>Введите простой чеклист для приемки моделей: <ul style="list-style-type: none"> <li>— при постановке задачи сразу фиксируется информация о таргете и выборке;</li> <li>— все задачи принимаются только через репозиторий, после обучения модель сохраняется в git, выборки — в базе данных, фиксируются метрики</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Заранее продумайте, как можно встроить ваш MVP в процесс принятия решений.</li> <li>Продумайте, кто и как будет делать ревью каждой компоненты.</li> <li>Тщательно продумайте тесты и минимально необходимые виды мониторинга для нового сервиса (мониторинг входящих данных, мониторинг результатов)</li> </ul>	<ul style="list-style-type: none"> <li>Продумайте, какие алерты и когда вы будете получать, как будете реагировать на них.</li> <li>Продумайте план действий при обновлении модели, выводе ее из эксплуатации.</li> <li>Продумайте план действий при изменении или добавлении источников данных (всякое бывает).</li> <li>Обеспечьте базовый контроль источников данных — и по наполняемости, и по содержанию (например, в поле были паспортные данные, а потом они стали приходить в каком-то другом формате). Причем это может произойти и с внутренними данными, если у вас большая экосистема, много продуктов и не один процессинг. Продумайте процесс миграции на другой тип данных.</li> <li>Для CI/CD можно использовать: <ul style="list-style-type: none"> <li>— систему патчсетирования для внедрения изменений в базу микросервиса;</li> </ul> </li> </ul>

скоринг \ data science \ разработка моделей

## Лучшие практики data science проектов для управления модельным риском: опыт QIWI

Окончание таблицы 2

Проверка гипотезы (MVP)	Deploy	Эксплуатация
(мы используем Jira с привязанным git и храним артефакты в Hadoop); — code review включает проверку заполнения данных о модели в каталоге, запуск модели на тестовых данных		— автоматизированные юнит- и интеграционные тесты; — отдельный сервис мониторинга для превентивного выявления проблем в данных или деградации моделей на ежедневной основе

моделей у нас весьма скромный, но поддержкой платформы занимается отдельное подразделение.

Особо хотелось бы подчеркнуть необходимость ревью модели, code review всех пайплайнов, потому что в целом внедрение моделей не слишком отличается от обычной продуктовой разработки — разработки backend или приложений. Все обычные практики программистов здесь полностью применимы:

- code review моделей, процедур поставки и трансформации данных, хотя бы самые простые репозитории в git;
- простые юнит-, интеграционные и нагрузочные тесты специально для DS-сервисов;
- отработанные процедуры «наката»/«отката» изменений;
- никаких изменений руками в проде — все делается через continuous integration и continuous delivery. Например, нам нравится использовать TeamCity, настраивать там автоматизированный flow — это действительно удобно.

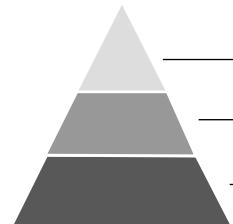
Большинство применяемых алгоритмов бустинговые, поэтому мы активно используем контейнеризацию в продакшн.

### Мониторинг управления модельным риском

Минимальный набор для мониторинга показан на рис. 1.

Рисунок 1

#### Минимальный набор для мониторинга



- Сбор метаданных, метрик и формирование автоматизированных контролей
- Логирование вектора фич и исходящих результатов
- Сохранение обучающей выборки, матрицы фич и результата

## Лидия ХРАМОВА

### Рекомендации относительно PSI

Для оценки изменений в модели можно сравнивать распределение скорбаллов на обучающей и текущей (на горизонте последних 1/30/90 дней) выборках с помощью известного индекса PSI:

$$\text{PSI} = \sum_{i=1}^N (y_{ti} - y_{bi})(\ln(y_{ti}) - \ln(y_{bi})) = \sum_{i=1}^N (y_{ti} - y_{bi}) \ln \frac{y_{ti}}{y_{bi}}.$$

PSI < 0,1 показывает отсутствие значимого изменения в текущей выборке.

PSI от 0,1 до 0,25 свидетельствует о незначительных изменениях, которые необходимо исследовать.

PSI > 0,25 говорит о значительном смещении популяции. Требуется перестроить модель.

Исключения из этого правила:

1) поменялся сегмент пользователей, соответственно изменились распределение входящих данных и распределение скорбалла;

2) возникли внешние обстоятельства, влияющие на кредитоспособность (локдаун, закрытие границ и т.д.).

При высоком PSI можно предпринять одно из следующих действий:

1. Выясните, поменялся ли сегмент пользователей (структура привлечения клиентов, структура каналов и пр.), изменились ли антифрод-политики или правила верификации.

2. Используя имеющиеся данные о просрочках в новом сегменте, проанализируйте, насколько хорошо ваша модель предсказывает дефолтность этого сегмента. Посчитайте на этих данных ROC AUC.

3. Проанализируйте изменения в распределениях фич. Они соответствуют новым сегментам или просто неправильно собираются?

Высокий PSI должен насторожить, если сегмент пользователей не поменялся, но распределение скорбалла стало другим. Скорее всего, модель придется переделать.

В качестве иллюстрации подхода, описанного в статье, продемонстрируем результаты ежедневного мониторинга в Grafana (рис. 2). На первом графике вверху PSI по каждой фиче, внизу данные для оценки пропусков (в реальности метрик в три раза больше: среднее, СКО и другие показатели). На втором графике результаты мониторинга скорбалла.

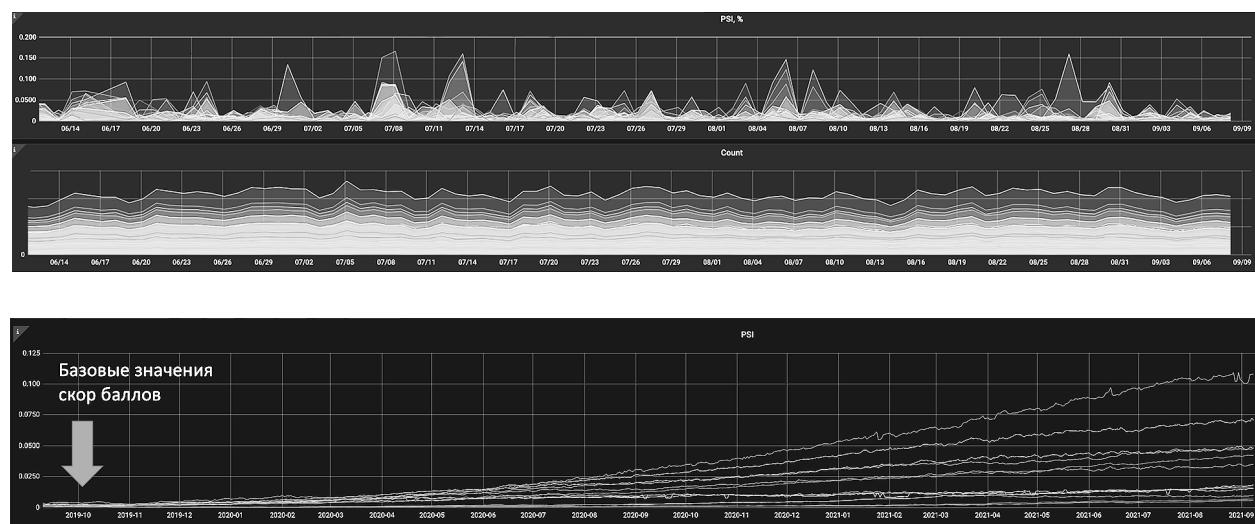
Мы фиксируем тестовую выборку для каждой модели в производстве, ежедневно делаем скринг большого количества отобранных людей каждой моделью и рассчитываем PSI. Данные октября 2019 г. — это базовая выборка.

скоринг \ data science \ разработка моделей

## Лучшие практики data science проектов для управления модельным риском: опыт QIWI

Рисунок 2

### Мониторинг фич и изменения скорбаллов



Таким образом, мы можем, еще не запустив модель в конвейер, оценить, сколько времени она проживет и насколько быстро деградирует. Life time наших моделей больше двух лет, и даже пандемия не повлияла на этот показатель. 

Европейская служба банковского надзора начинает практику сближения полноценных ML-моделей с регулятивными требованиями к расчету риска на капитал. В связи с тем, что Россия часто впереди планеты всей в имплементации Базеля и продвинутых практик регулирования, банковским датасаентистам, возможно, уже сейчас стоит озабочиться подготовкой к таким изменениям.

## Полноценный ML в моделях IRB: прощай, логрег?

Европейская служба банковского надзора, известная своими монументальными подходами к банковским моделям (см., напр., Guidelines on PD estimation, LGD estimation and the treatment of defaulted exposures<sup>1</sup> — 200 страниц про то, как строить модель логистической регрессии), выпустила знаменательный документ<sup>2</sup>. В нем она предлагает финансовым кругам обсудить достоинства «полноценных» ML-моделей, в которые включены «Tree, Generalised regression, Neural network, Boosting». Это может быть началом эпохи «больших» моделей в регулировании, которые в контексте требований на капитал ранее использовались только для отбора признаков в IRB-модели или в задачах, не связанных с регулированием вообще.

### Что не так с моделями IRB

ЕВА отмечает, что модели на основе внутренних рейтингов (IRB), используемые для расчета нормативных требований к капиталу в отношении кредитного риска, существенно не отличаются от подходов, применявшимися 15–20 лет назад, когда было введено в действие первое Базельское соглашение. С тех пор внимание регуляторов и надзорных органов было сосредоточено больше на том, чтобы сделать оценки, полученные с помощью различных моделей, сопоставимыми путем улучшения определений основных понятий (например, дефолта), чем на понимании проблем передовых технологий — машинного обучения (ML) и искусственного интеллекта (AI).

<sup>1</sup> EBA/GL/2017/16.

<sup>2</sup> EBA discussion paper on machine learning for IRB models. EBA/DP/2021/04. 11 November 2021.

## Полноценный ML в моделях IRB: прощай, логрэг?

Модели ML могут быть полезны для повышения предсказательной силы внутренних моделей и не являются чем-то новым, но они не были включены в модели IRB так быстро, как в других областях. Основная проблема связана с их сложностью, что приводит к проблемам в (1) интерпретации их результатов, (2) обеспечении их адекватного понимания руководством кредитных учреждений и (3) обосновании их результатов перед надзорными органами.

Такой шаг навстречу здравому смыслу наверняка будет сопровождаться тяжелыми инфраструктурными требованиями.

### Что нас ожидает

Основные рекомендации приведены в таблице.

Таблица

### Рекомендации ЕВА

Что придется делать	Чего нужно избегать	Какие проблемы могут возникнуть
<p>1. Анализировать статистическими методами:</p> <ul style="list-style-type: none"> <li>— связь каждого риск-фактора с выходной переменной при прочих равных условиях;</li> <li>— вес каждого риск-фактора.</li> </ul> <p>2. Оценивать экономическую связь каждого риск-фактора с выходной переменной, чтобы убедиться, что оценки модели правдоподобны и интуитивно понятны.</p> <p>3. Если при разработке модели используется экспертное суждение — оценивать, вносят ли выбранные риск-факторы вклад в оценку риска в соответствии с их экономическим смыслом.</p> <p>Если экспертное суждение используется при применении модели — сотрудники, отвечающие за переопределение результатов модели, должны быть в состоянии рассмотреть поведение модели на конкретном прогнозе, принимая во внимание аспекты, которые еще не заложены в нее, или определить случаи, когда логика модели может ввести в заблуждение.</p> <p>Доверять переопределение результатов автоматизированной модели, учитывая только те аспекты, которые недостаточно встроены в автоматический рейтинг.</p> <p>4. Если ML-модели часто обновляются — анализировать причину, т.к. кредитный риск (в отличие от рыночного) не подвержен быстрым изменениям. Оправдать обновление модели могут изменение экономических условий, процессов в банке или базовых данных.</p>	<p>1. Чрезмерного количества объясняющих переменных или переменных без значимой прогностической информации.</p> <p>2. Использования неструктурированных данных, если имеются более традиционные данные, обеспечивающие аналогичные возможности прогнозирования.</p> <p>3. Чрезмерно сложных вариантов моделирования при наличии более простых подходов, дающих аналогичные результаты</p>	<p>1. Чрезмерная подгонка (оверфиттинг).</p> <p>2. Влияние экспертного мнения на гиперпараметры, используемые для описания структуры модели и настройки алгоритма обучения. Отдел валидации должен уделять особое внимание проверке обоснования выбора этих гиперпараметров. Если гиперпараметры выбираются путем минимизации ошибки модели, необходимо убедиться, что этот процесс не вносит нежелательного смещения.</p> <p>3. Проблемы презентативности и качества данных. Если данные внешние, рекомендуется уделять особое внимание оценке их презентативности.</p> <p>4. Проблема стабильности оценок, в том числе с учетом рейтинговой философии учреждения. Алгоритмы ML могут вводить в модель элементы point-in-time, способные нарушить стабильность процесса присвоения рейтинга по сравнению с моделями through-the-cycle, ведущими к быстрым изменениям требований к капиталу</p>



## Окончание таблицы

Что придется делать	Чего нужно избегать	Какие проблемы могут возникнуть
<p>Сравнивать изменения с последней утвержденной моделью, чтобы убедиться, что множество незначительных частых изменений не приведут к незамеченному существенному изменению.</p> <p>5. Подготовить краткий документ для всех заинтересованных сторон, где модель объясняется в простой форме на основе результатов п. 1. В документе должны быть описаны как минимум:</p> <ul style="list-style-type: none"> <li>— ключевые драйверы модели;</li> <li>— основные взаимосвязи между риск-факторами и прогнозами модели.</li> </ul> <p>6. Организовать углубленные тренинги для сотрудников отделов разработки моделей, контроля кредитного риска и валидации.</p> <p>7. Обеспечить предоставление органу управления и высшему руководству документации, разъясняющей, какие показатели/переменные являются ключевыми и (если уместно) как ML-модели влияют на количественную оценку риска</p>		

## Примеры

Создадим простую бустинговую модель на данных соревнования FICO explainable-machine-learning-challenge<sup>1</sup> упрощенным кодом<sup>2</sup> вроде:

```
import xgboost as xgb
params = {
    "max_depth": 5,
    "objective": "binary:logistic",
    "eval_metric": "auc"
}

train = xgb.DMatrix(X, label=y)
classifier = xgb.train(params, train, verbose_eval=1)
```

Пробежимся по техническим заметкам документа, который говорит: «The use of these techniques can also pose a challenge by itself in relation to the mathematical hypotheses on which they rely, the difficulty to implement them or due to the computational capacities required». Однако для математических сложностей как будто создана библио-

<sup>1</sup> <https://community.fico.com/s/explainable-machine-learning-challenge>.

<sup>2</sup> Здесь и далее мы не приводим весь код, а даем выдержки, показывающие концепцию на Python.

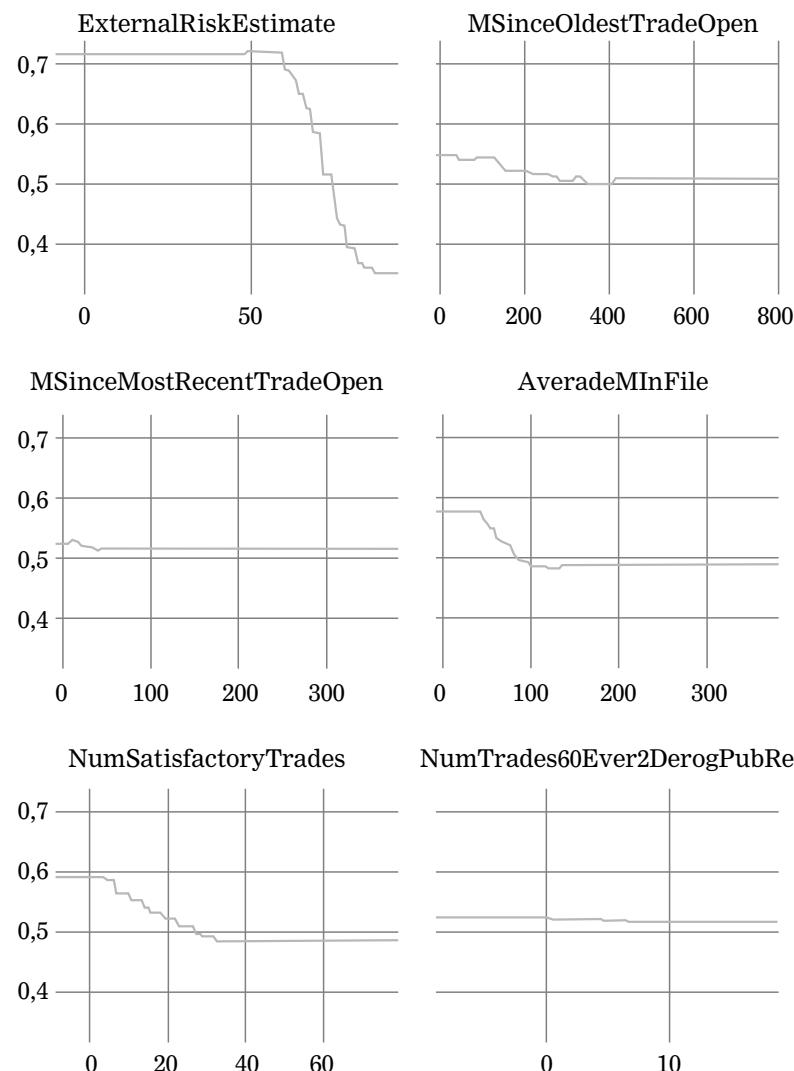
## Полноценный ML в моделях IRB: прощай, логрэг?

тека dalex<sup>1</sup>; более того, как будет видно далее, ее методы прямо имплементируют положения документа ЕВА — или наоборот, учтывая время создания библиотеки (2018 г.).

1. Graphical tools showing the effect of an explanatory variable on the model.

График создается упрощенным кодом, например:

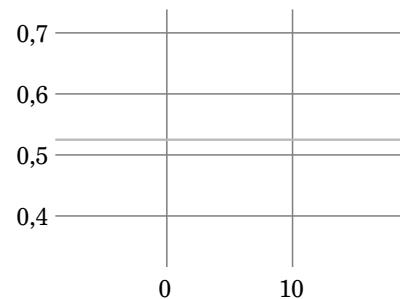
```
exp = dx.Explainer(classifier, X, y)
pd_rf = exp.model_profile()#(variable = "target", type = "partial")
pd_rf.plot()
```



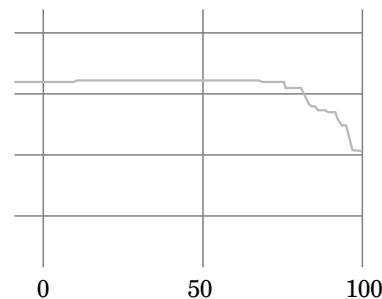
<sup>1</sup> <https://github.com/ModelOriented/DALEX>.



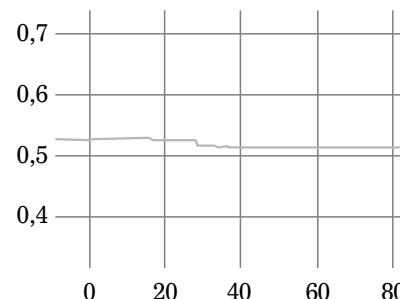
NumTrades90Ever2DerogPubRec



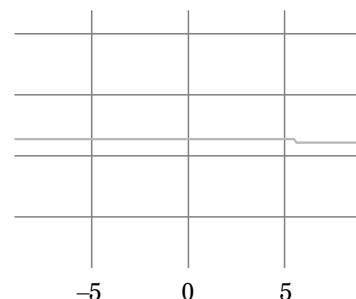
PercentTradesNeverDelq



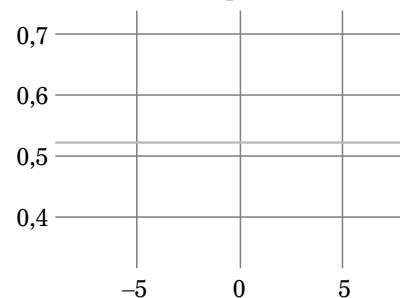
MSinceMostRecentDelq



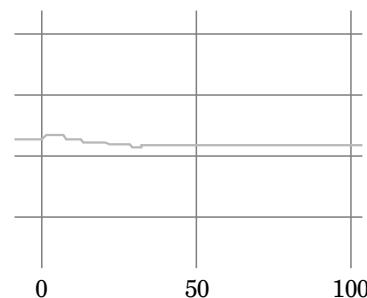
MaxDelq2PublicRecLast12M



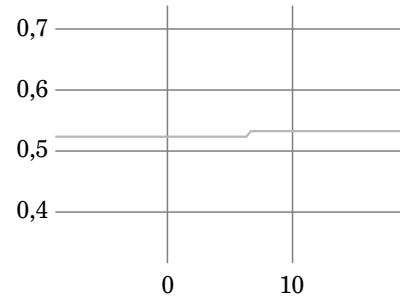
MaxDelqEver



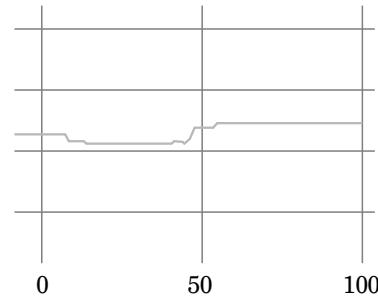
NumTotalTrades



NumTradesOpeninLast12M

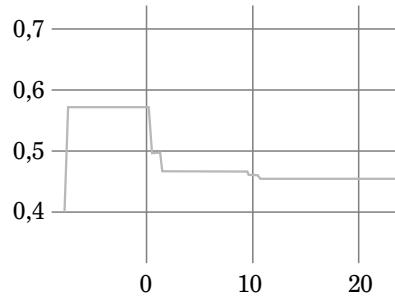


PercentInstallTrades

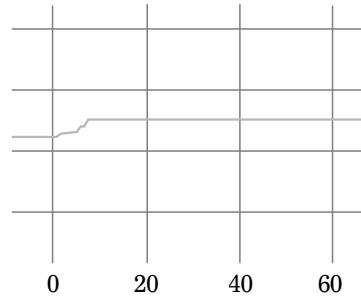


## Полноценный ML в моделях IRB: прощай, логрэг?

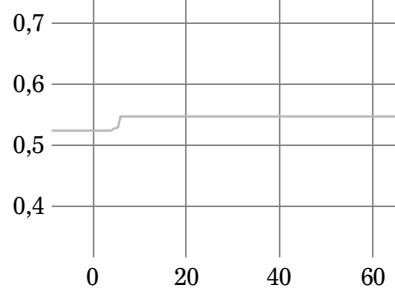
MSinceMostRecentInqexcl7days



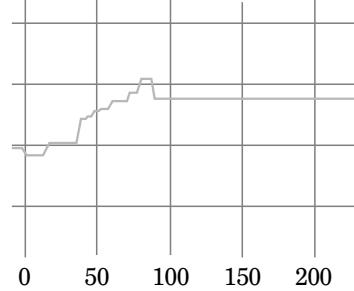
NumInqLast6M



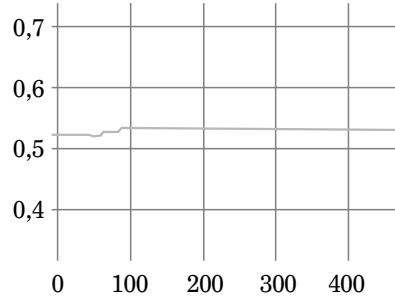
NumInqLast6Mexcl7days



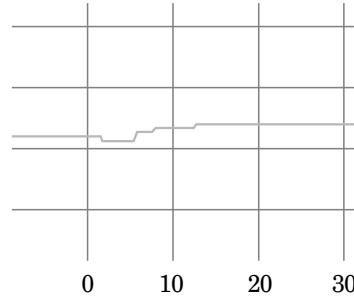
NetFractionRevolvingBurden



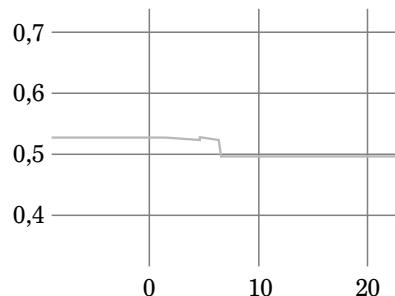
NetFractionInstallBurden



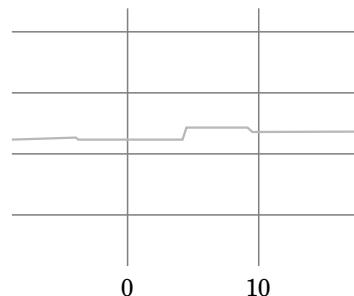
NumRevolvingTradesWBalance



NumInstallTradesWBalance

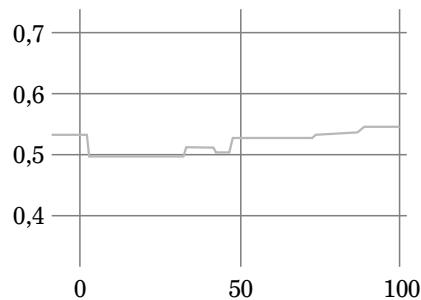


NumBank2NatlTradesWHighUtilization





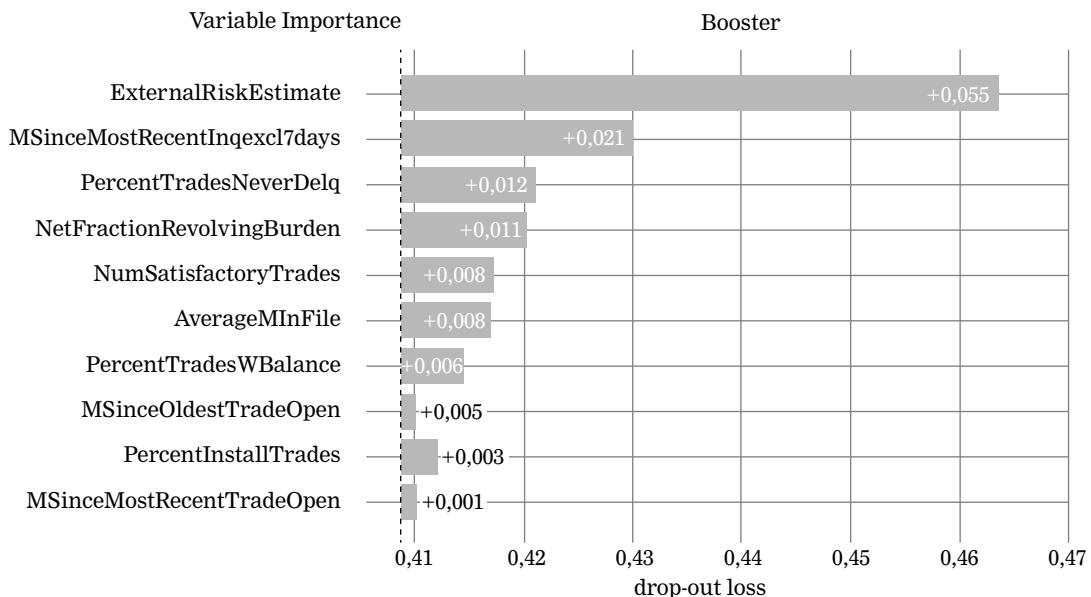
PercentTradesWBalance



2. Feature importance measures reveal the relevance of each explanatory variable in the overall model.

График создается упрощенным кодом вроде:

```
exp.model_parts().plot()
```

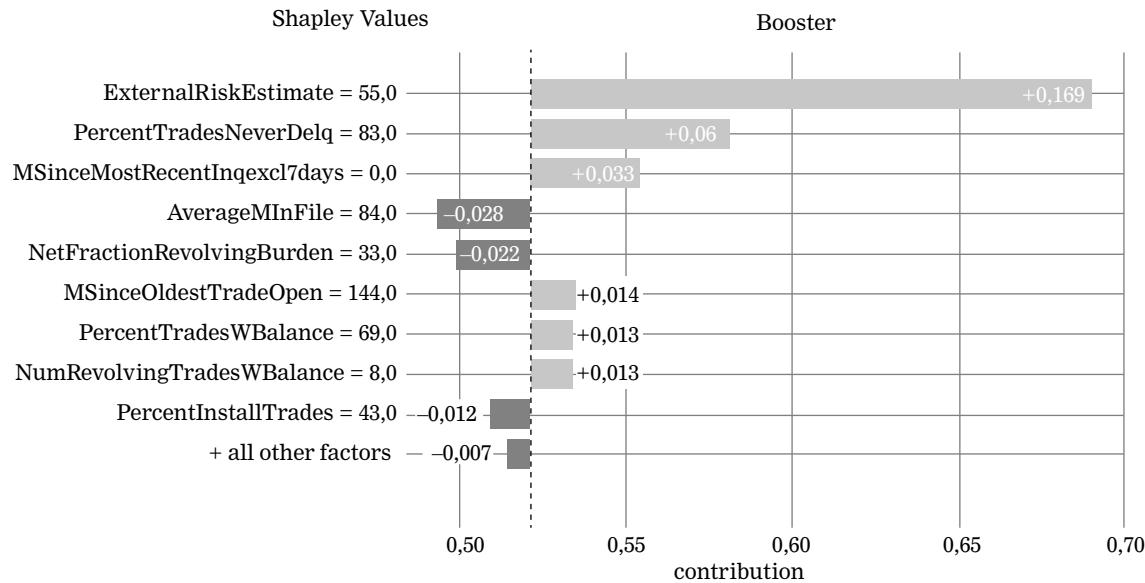


3. Shapley values quantify the impact of each explanatory variable on a specific prediction of the model.

График создается упрощенным кодом вроде:

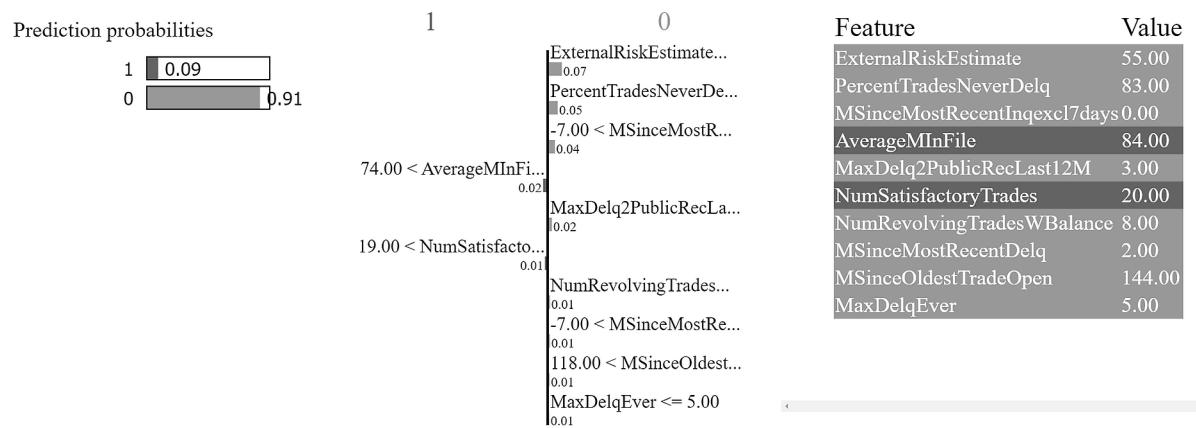
```
bd = exp.predict_parts(X.iloc[0,:], type = 'shap')
bd.plot()
```

## Полноценный ML в моделях IRB: прощай, логрэг?



4. Local explanations, such as LIME and anchors, provide simple approximations of the model on a vicinity of an observation.

Здесь, к сожалению, используется иная библиотека — Lime, которая требует оберток вокруг xgboost-модели и генерирует немного архаичные, но вполне понятные графики:

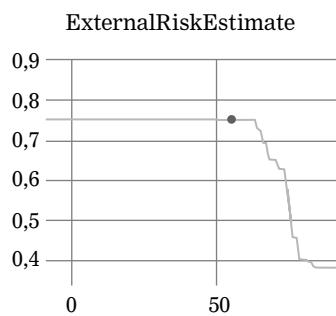


5. Counterfactual explanations indicate how a specific prediction of the model could be modified by altering the values of the explanatory variables as little as possible.

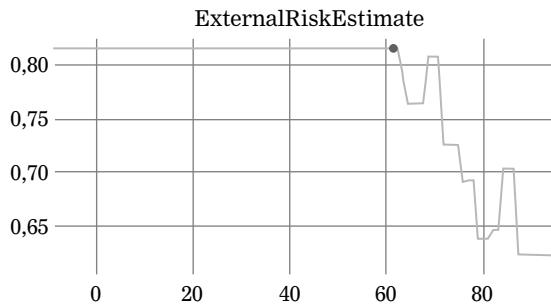
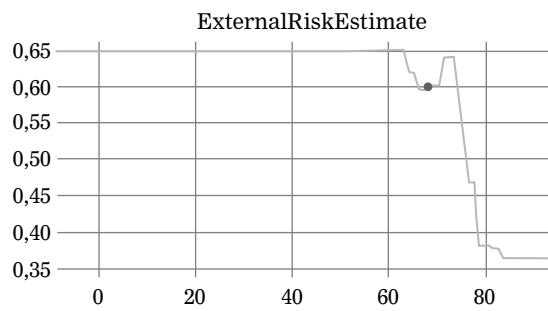
Вернемся в библиотеку dalex и получим матрицу ceteris paribus упрощенным кодом вида:

```
cp = exp.predict_profile(X.iloc[0,:])
cp.result
```

На примере одного наблюдения мы можем понять, как ведет себя модель, и построить следующий график, иллюстрирующий изменение предсказания в зависимости от признака ExternalRiskEstimate (согласно определениям датасета — предиктор, консолидированный риск-маркер):



Однако надо помнить, что это не объяснение модели, и нельзя сказать, что предсказание монотонно убывает при увеличении значения данного предиктора. Приведем примеры этого графика на других точках данных:



---

## Полноценный ML в моделях IRB: прощай, логрэ?

---

В связи с этим вполне объяснимой выглядит техническая аннотация: «Also, it must be noted that each technique provides only some partial understanding of a model, and that their usefulness can greatly vary depending on the case». Действительно, ряд методов может ввести в заблуждение, а необходимые технические оговорки могут значительно снизить их ценность.

Возможно, самым действенным объяснением особенностей применяемых «сложных» моделей для стейкхолдеров будет табличное объяснение, как работают дерево решений и градиентный спуск. Для руководителей понимание этих вопросов, похоже, становится потребностью уже сегодняшнего дня. 

## Эмилио ЛЬОРЕНТЕ

В крупных банках мнение инвестиционного комитета все еще превалирует над данными скрупулезного анализа рыночной ситуации и моделями, построенными на них. Разработанная автором на основе MATLAB система принятия решений по активам призвана заменить решения комитета data-driven решениями<sup>1</sup>. Для этого члены команды RAMS<sup>2</sup> внимательно изучили мнение каждого эксперта, заменили его моделью и снабдили текущими рыночными данными на основании Мerton-моделей. Заметят ли члены комитета разницу?

Эмилио ЛЬОРЕНТЕ, *Recognition Asset Management Solutions*, генеральный директор

# Как с помощью MATLAB разместить деньги в активы и снизить рыночные и операционные риски

## Динамическое управление активами

Динамическое распределение активов без созыва комитета позволяет снизить операционные и рыночные риски за счет адаптации решений, близких к человеческим, к событиям на рынке.

Возьмем некий стратегический портфель клиента (банковского учреждения), в котором подытожим анализ аллокаций капитала клиента в классы активов (таблица).

В рамках различных сценариев цифры сильно отличаются. Какой профиль размещения средств выбрать, чтобы максимально соответствовать «человеческому» комитету, и какую методологию расчетов использовать?

## Как мыслит эксперт

Эксперт в ракурсе принятия инвестиционных решений мыслит в пределах лимитов, текущей рыночной ситуации и собственных знаний.

Мы обобщили это как некий многоокритериальный подход к оценке нескольких активов согласно рис. 1.

<sup>1</sup> Статья подготовлена на основе выступления автора на конференции MATLAB Expo 2021: <https://www.matlabexpo.com/>.

<sup>2</sup> Recognition Asset Management Solutions.

## Как с помощью MATLAB разместить деньги в активы и снизить рыночные и операционные риски

Таблица

### Стратегический портфель

	Minima (%)	Neutral (%)	Maxima (%)
<b>Equities</b>	<b>18,0</b>	<b>36,0</b>	<b>50,0</b>
Equity US	8,0	14,0	25,0
Equity Europe	8,0	14,0	25,0
Equity Japan	1,0	3,5	6,0
Equity GEM	1,0	4,5	8,0
<b>Government</b>	<b>9,0</b>	<b>20,5</b>	<b>32,0</b>
US Treasury	3,0	7,5	12,0
Eurozone Government	6,0	13,0	20,0
<b>Credit</b>	<b>16,0</b>	<b>34,0</b>	<b>52,0</b>
Investment Grade US	6,0	10,5	15,0
Investment Grade Eurozone	8,0	11,5	15,0
High Yield US	1,0	4,5	8,0
High Yield Europe	1,0	4,5	8,0
GEM Credit	0,0	3,0	6,0
<b>Alternative</b>	<b>0,0</b>	<b>8,0</b>	<b>21,0</b>
<b>Cash</b>	<b>0,0</b>	<b>1,5</b>	<b>3,0</b>

Рисунок 1

### Процесс принятия решений экспертом

- Analyst: Global Asset Allocation Expert A
- Focus: Top Asset Classes
- Granularity: Level 1

Decision survey							
Global Assets							
Criteria:	Macroeconomics	Sentiment	Policy	Value	Technical	Systemic_Risk	
Importance:	High	Medium	Medium_High	High	Medium	Medium_Low	
EQUITIES	Overweight	Overweight	Overweight	Neutral	Mild_Overweight	Mild_Underweight	
GOVERNMENT	Underweight	Underweight	Neutral	Strong_Underweight	Underweight	Mild_Overweight	
CREDIT	Mild_Overweight	Mild_Overweight	Neutral	Mild_Underweight	Mild_Overweight	Mild_Overweight	
ALTERNATIVE	Mild_Overweight	Overweight	Overweight	Overweight	Mild_Underweight	Neutral	
CASH	Mild_Overweight	Mild_Underweight	Neutral	Underweight	Mild_Overweight	Mild_Overweight	

### Методология моделирования экспертного мнения

Мы используем два основных метода на основе MATLAB. Первый — это нечеткая логика, а второй — TOPSIS. Нечеткая логика позволит преобразовать слова, мнения и критерии в числа, а TOPSIS — объединить

## Эмилио ЛЬОРЕНТЕ

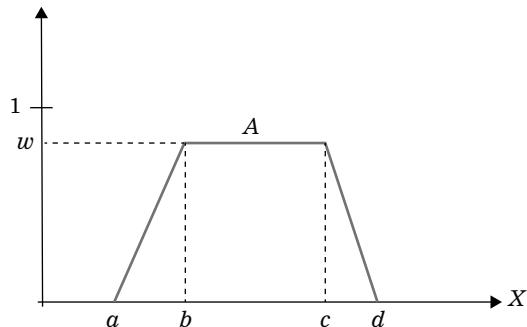
эти наборы чисел как подходы экспертов с одним-единственным мнением, которое должно исходить от инвестиционного комитета.

Нечеткая логика — это перспективная теория, которая может рассматриваться как расширение классической теории множеств.

Представим обобщенное трапециевидное число  $A$  как  $A = (a, b, c, d; w)$ , где  $a, b, c, d$  — реальные величины,  $0 < w \leq 1$  (рис. 2).

Рисунок 2

### Отображение данных в рамках нечеткой логики



Предположим, что у нас есть два обобщенных трапециевидных числа  $A$  и  $B$ , где  $A = (a_1, a_2, a_3, a_4, w_a)$  и  $B = (b_1, b_2, b_3, b_4, w_b)$ . Степень сходства между этими обобщенными числами может быть рассчитана следующим образом:

$$S(\tilde{A}, \tilde{B}) = \left( 1 - \frac{\sum_{i=1}^4 |a_i - b_i|}{4} \right) \times \frac{\min(P(\tilde{A}), P(\tilde{B})) + \min(w_a, w_b)}{\max(P(\tilde{A}), P(\tilde{B})) + \max(w_a, w_b)}$$

MATLAB позволяет преобразовать мнения в нечеткие числа, числа — в соглашение между членами комитета. Появляется возможность ранжировать конкретные цифры, которые позволят разместить активы банка (рис. 3).

Далее с помощью Bloomberg-коннектора мы получаем расширенные данные по всем классам активов.

Данные получены из 5 экономических регионов по более чем 1000 компаний за последние 15 лет, примерно по 7 стоимостным и рыночным коэффициентам. Информация максимально детализирована и включает индексы, инфляционные ожидания, спред, подразумеваемую волатильность, distance-to-default по Мертон-модели.

## Как с помощью MATLAB разместить деньги в активы и снизить рыночные и операционные риски

Рисунок 3

### Выходные данные модели

Equities	 0.571	Neutral-Overweight
		Equity US  0.507 Neutral
		Equity Europe  0.429 Neutral-Negative
		Equity Japan  0.404 Neutral-Negative
		Equity GEM  0.614 Positive
Government	 0.349	Underweight
		US Treasury  0.524 Neutral-Positive
		Eurozone Government  0.476 Neutral-Negative
Credit	 0.512	Neutral
		Investment Grade US  0.504 Neutral
		Investment Grade Eurozone  0.412 Neutral-Negative
		High Yield US  0.410 Neutral-Negative
		High Yield Europe  0.390 Negative
		GEM Credit  0.644 Positive
		GEM Credit HC  0.544 Neutral-Positive
		GEM Credit LC  0.457 Neutral-Negative
Alternative	 0.617	Overweight
		HF Equally Weighted  0.492 Neutral
		Market Neutral  0.575 Neutral-Positive
		Commodities  0.440 Neutral-Negative
Cash	 0.482	Neutral-Underweight

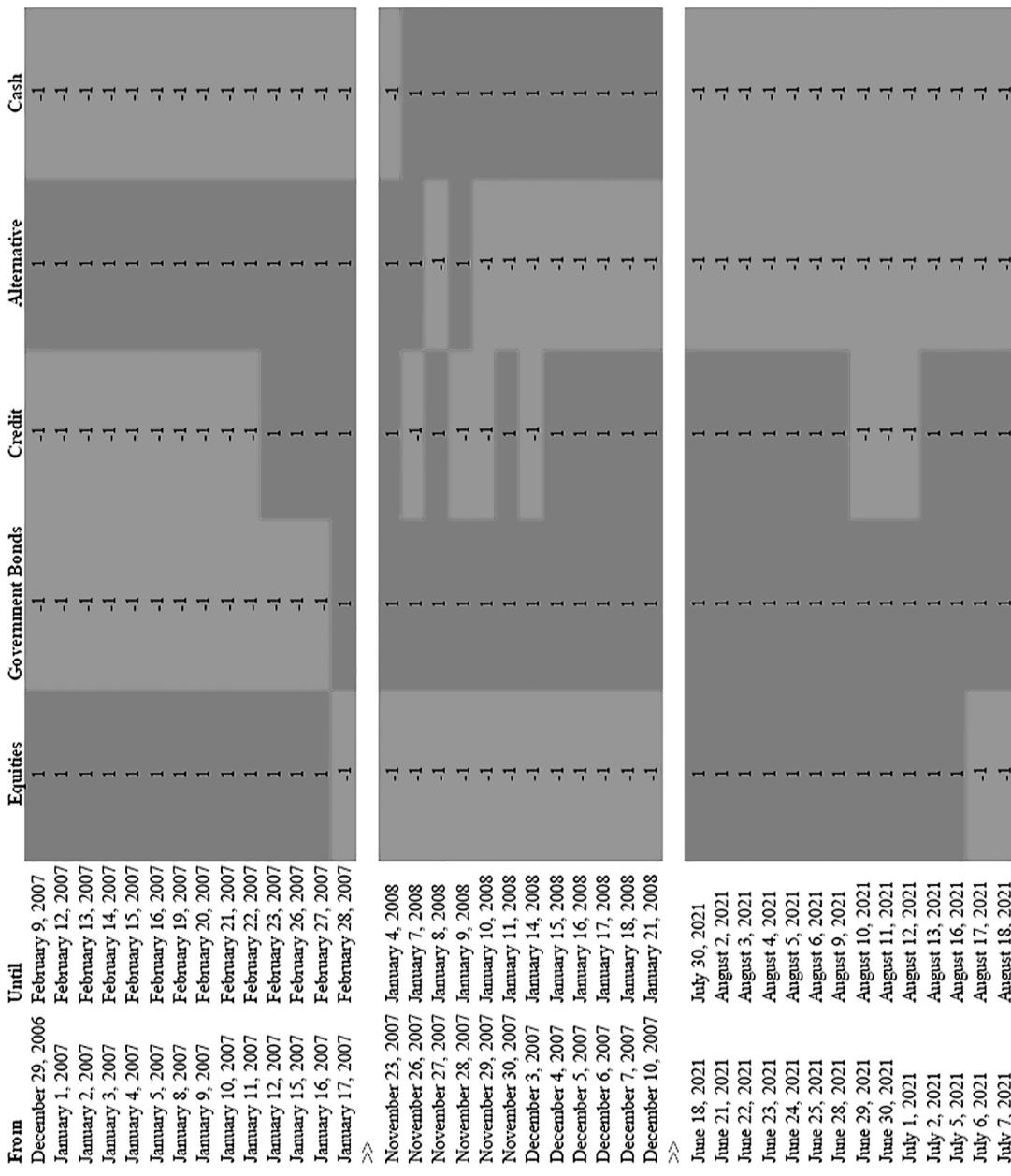
Концепция Мертона позволяет определить логичные и интерпретируемые сценарии: предполагаемые рынком вероятности дефолта являются циклическими и сильно коррелируют с реализованными коэффициентами дефолта, повышаются в периоды чрезмерного экономического кризиса и снижаются во время экономического подъема.

Данные поступают в следующую обученную модель на основе метода опорных векторов. Цель модели — спрогнозировать рост или падение каждого вида активов, а также его справедливую стоимость.

На рис. 4 представлен результат для активов первого уровня, который мы ищем, пытаясь выявить предпочтительные классы активов для того тактического временного горизонта, который мы определили с клиентом. Мы должны построить процесс машинного обучения для ранжирования активов, аналогичного ранжированию, которое делают эксперты. Таким образом мы получаем многофункциональные модели.

## ЭМИЛИО ЛЬОРЕНТЕ

### Результат для различных категорий активов первого уровня



## Как с помощью MATLAB разместить деньги в активы и снизить рыночные и операционные риски

Рисунок 5

### Ранжирование человеком и машиной: сравнение

- Human experts Investment Committee, TAA Ranking
- Focus: All Asset Classes
- Granularity: Level 1, 2, 3

Equities	0.571	Neutral-Overweight
		Equity US  0.507 Neutral
		Equity Europe  0.429 Neutral-Negative
		Equity Japan  0.404 Neutral-Negative
		Equity GEM  0.614 Positive
Government	0.349	Underweight
		US Treasury  0.524 Neutral-Positive
		Eurozone Government  0.476 Neutral-Negative
Credit	0.512	Neutral
		Investment Grade US  0.504 Neutral
		Investment Grade Eurozone  0.412 Neutral-Negative
		High Yield US  0.410 Neutral-Negative
		High Yield Europe  0.390 Negative
		GEM Credit  0.644 Positive
		GEM Credit HC  0.544 Neutral-Positive
		GEM Credit LC  0.457 Neutral-Negative
Alternative	0.617	Overweight
		HF Equally Weighted  0.492 Neutral
		Market Neutral  0.575 Neutral-Positive
		Commodities  0.440 Neutral-Negative
Cash	0.482	Neutral-Underweight

- Machine Learning, TAA Ranking
- Focus: All Asset Classes
- Granularity: Level 1, 2, 3

Equities	0.646	Overweight
		Equity US  0.805 Very Positive
		Equity Europe  0.458 Neutral-Negative
		Equity Japan  0.538 Neutral-Positive
		Equity GEM  0.462 Neutral-Negative
Government	0.024	Strong Underweight
		US Treasury  0.532 Neutral-Positive
		Eurozone Government  0.468 Neutral-Negative
Credit	0.718	Overweight
		Investment Grade US  0.045 Very Negative
		Investment Grade Eurozone  0.238 Negative
		High Yield US  0.950 Very Positive
		High Yield Europe  0.950 Very Positive
		GEM Credit  0.388 Negative
		GEM Credit HC  0.812 Very Positive
		GEM Credit LC  0.345 Negative
Alternative	0.745	Overweight
		HF Equally Weighted  0.424 Neutral-Negative
		Market Neutral  0.680 Positive
		Commodities  0.356 Negative
Cash	0.379	Underweight

## Эмилио ЛЬОРЕНТЕ

На рис. 5 показана разница между машинным «комитетом» и человеческим.

Мы видим, что за счет нечеткой логики и метода опорных векторов удалось добиться значительного сходства между двумя классами систем принятия решений.

Рыночные риски будут снижены за счет учета максимального количества данных, операционные — за счет минимального участия человека в подобных «комитетах». Возможно, такая модель будет актуальна для цифровых банков будущего. 

Как совместить риск-нейтральную меру ARIMA-GARCH с необходимостью учета тяжелых хвостов, характерных для распределений финансовых временных рядов? В статье на практическом примере показана действующая модификация расширенного принципа Гирсанова, в котором вместо логарифмических приращений берутся относительные, и приведены примеры кода в среде R.

## Численные эксперименты оценки VaR/CVaR портфеля опционных контрактов: новейшие подходы

Наша задача — оценить меры риска VaR/CVaR портфеля из 15 опционных контрактов на 15 различных базовых активов (таблица).

В качестве входных параметров алгоритма выступают (в скобках указаны значения, взятые в качестве примера):

- количество сценариев численного метода Монте-Карло (10 000 (MonteVaR));
- процент исходной дисперсии, который будет описываться методом главных компонент (96%);
- распределение ошибки моделей ARIMA-GARCH (нормальное);
- уровень надежности (95%);
- период расчета (10 дней).

Для расчета мер риска VaR/CVaR методом Монте-Карло мы:

1. Из исходных цен  $S_t$  базовых активов получили их доходности

$$\tilde{Y}_t^j = \frac{S_t}{S_{t-1}} - 1.$$

2. Применили метод главных компонент и перешли от доходностей цен базовых активов  $\tilde{Y}_t^j$  к рядам компонент  $X_t^i$  через матрицу перехода  $a_j^i$ .

`ept.pca = prcomp(returns, scale. = FALSE)`

3. Выбирая уровень дисперсии, который будет описывать компоненты исходных доходностей, сократили размерность исходной задачи с  $n$  до  $m$ .



Артем  
ДАНИЛИШИН,  
Промсвязьбанк,  
Управляющий риск-  
менеджер Группы  
моделирования  
и внедрения продуктов

## Артем ДАНИЛИШИН

Таблица

### Состав портфеля опционных контрактов

Базовый актив	Цена исполнения	Дата исполнения	Тип контракта
HEQ1	104	20.08.2021	Call
ZMQ1	360	20.08.2021	Put
LEQ1	120	20.08.2021	Call
NGQ1	4	20.08.2021	Put
CLU1	71,5	17.09.2021	Call
BZU1	74	17.09.2021	Call
HOQ1	2,13	20.08.2021	Call
6CU1	0,81	17.09.2021	Call
6EU1	1,165	17.09.2021	Put
6BU1	1,38	17.09.2021	Call
6AU1	0,74	17.09.2021	Call
6JU1	0,00895	17.09.2021	Put
GCV1	1800	15.10.2021	Call
HGU1	4,4	17.09.2021	Call
SIU1	25,25	17.09.2021	Call

nComp = which(cumsum(hept.pca\$sdev^2 / sum(hept.pca\$sdev^2)) > DispComp)[1]

4. Оценили ARIMA-GARCH модели первых  $m$  главных компонент.

```
fit = list(dim = nComp)
for(i in 1:nComp){
  fit[[i]] = ugarchfit(spec = spec,data= hept.pca$x[,i],solver = 'solnp')
}
```

5. Сгенерировали сценарии (для каждой компоненты)  $m_t^l$ ,  $h_t^l$ ,  $\varepsilon_t^l$ , где  $l = \overline{1:M}$  — номер сценария.

```
for (i in 1:nComp){
  ttt = ugarchsim(fit[[i]], n.sim = dim(simul)[2], n.start = 0, m.sim = nmonte,
  startMethod= "sample")
  simul[1,,i,] = ttt@simulation$residSim
  simul[2,,i,] = ttt@simulation$seriesSim - ttt@simulation$residSim
}
```

---

## Численные эксперименты оценки VaR/CVaR портфеля опционных контрактов: новейшие подходы

---

6. На основе модификации расширенного принципа Гирсанова совместно с методом главных компонент получили подход, позволяющий моделировать совместную риск-нейтральную динамику случайных процессов. Данный подход применен к оцениванию мер риска VaR/CVaR портфеля опционных контрактов:

$$\tilde{Y}_t^{jl} = \left(1 + \frac{r_j}{n}\right)^n - 1 + \left( \frac{\left(1 + \frac{r_j}{n}\right)^n}{1 + \sum_{i=1}^m a_j^i m_t^{il}} \right) \sqrt{\sum_{i=1}^m a_j^{i2} h_t^{il}} E_t^{jl}, E_t^{jl} = \sum_{i=1}^m \frac{a_j^i \sqrt{h_t^{il}} \varepsilon_t^{il}}{\sqrt{\sum_{i=1}^m a_j^{i2} h_t^{il}}}.$$

```
out_assets_rn = array(data = 0, dim = c(max(days), nAssets, nmonte))
out_assets_ph = array(data = 0, dim = c(max(days), nAssets, nmonte))
for(i in 1:nAssets){
  temp_1 = array(data = 1, dim = c(max(days), nmonte))
  temp_2 = array(data = 0, dim = c(max(days), nmonte))
  temp_3 = array(data = 0, dim = c(max(days), nmonte))
  for(j in 1:nComp){
    temp_1 = temp_1 + hept.pca$rotation[i,j]*simul[2,,j]
    temp_2 = temp_2 + hept.pca$rotation[i,j]*simul[1,,j]
    temp_3 = temp_3 + hept.pca$rotation[i,j]*(simul[2,,j] + simul[1,,j])
  }
  out_assets_rn[,i,] = ((1+r[i]/nd)^^(nd)) - 1 + (((1+r[i]/nd)^^(nd))/(temp_1))*temp_2
  out_assets_ph[,i,] = temp_3
}
```

7. По  $\tilde{Y}_t^{jl}$  получили сценарии цен в момент времени экспирации опциона  $\tilde{S}_T^{jl}$ , а также значение функции выплаты ( $b_c(S_T, X) = \max(S_T - X, 0)$ ,  $b_p(S_T, X) = \max(X - S_T, 0)$ ).

```
mass_of_final_assets_rn = array(data = 0, dim = c(nAssets, nmonte))
mass_of_final_assets_ph = array(data = 0, dim = c(nAssets, nmonte))
for(q in 1:nAssets){
  for(mc in 1:nmonte){
    mass_of_final_assets_rn[q,mc] = prod(out_assets_rn[1:days[q],q,mc]+1)*as.double(datas[dim(datas)[1],1+q])
    mass_of_final_assets_ph[q,mc] = prod(out_assets_ph[1:days[q],q,mc]+1)*as.double(datas[dim(datas)[1],1+q])
  }
  if(datas[1,(nAssets+1+q)]=="CALL"){
    mass_of_final_assets_rn[q,] = pmax(mass_of_final_assets_rn[q,] - as.double(datas[2,(nAssets+1+q)]),0)
  }
}
```

## Артем ДАНИЛИШИН

```
mass_of_final_assets_ph[q,] = pmax(mass_of_final_assets_ph[q,] - as.double
(data$[2,(nAssets+1+q)]),0)
}else{
  mass_of_final_assets_rn[q,] = pmax(-mass_of_final_assets_rn[q,] + as.double
(data$[2,(nAssets+1+q)]),0)
  mass_of_final_assets_ph[q,] = pmax(-mass_of_final_assets_ph[q,] + as.double
(data$[2,(nAssets+1+q)]),0)
}
}
```

8. Добавили в массив сценариев (ScenariosOpt) приведенное среднее результата вычисления шагов 1–7.

```
for(i in 1:nAssets){
  ScenariosOpt[v] = ScenariosOpt[v] + sum(mass_of_final_assets_rn[i,] /
((1+r[i])^(days[i] - TVaR))) / nmonte
}
```

9. Повторили шаги 1–8 MonteVaR = 10 000 раз для получения статистики.

10. Рассчитали значения мер риска VaR/CVaR, используя значения массива ScenariosOp.

```
VaR = sort(ScenariosOpt)[as.integer(length(ScenariosOpt) * alpha + 1)]
CVaR = mean(sort(ScenariosOpt)[1:as.integer(length(ScenariosOpt) * alpha + 1)])
```

Описанный подход позволяет получать коэффициенты ARIMA-GARCH моделей на основе физической и риск-нейтральной меры, оценивать справедливую стоимость опционных контрактов методом Монте-Карло и на основе найденных стоимостей строить оценку VaR портфеля опционных контрактов. 

В КНР экономисты и датасаентисты, похоже, наконец пришли к согласию. Они выработали методологию и тип моделей, которые, во-первых, дружат между собой, а во-вторых, не встречаются ни на Github, ни на Kaggle. Попробуем посмотреть, в чем состоял смысл многолетней исследовательской работы в области кредитных рисков и анализа данных<sup>1</sup>.

## Исследование кредитных рейтингов компаний КНР: качество и интерпретабельность

В качестве исходных данных были взяты групповые данные китайских компаний, зарегистрированных на бирже, за период с 2000 по 2017 гг., полученные из базы данных Wind Economic Database.

Сначала осуществляется выбор необходимых признаков по так называемому индексу достоверности, рассчитываемому следующим образом.

Предположим, что набор данных  $S$  состоит из  $X_1, X_2, \dots, X_n$  и разделен на  $K$  количество классов  $C_1, C_2, \dots, C_k$ ;  $m_k$  обозначает количество наблюдений в  $C_k$ , а  $X_i^k$  — это  $i$ -е наблюдение в  $C_k$ . Центроид каждого класса выражен как  $r_0^1, r_0^2, \dots, r_0^K$  и определяется как:

$$r_0^k = \sum_{i=1}^{m_k} \frac{X_i^k}{m_k}. \quad (1)$$

Индекс внутриклассовой компактности определяется как:

$$\text{Intra}(h) = \frac{1}{K} \sum_{k=1}^K \text{Inner}(k). \quad (2)$$

Член  $\text{Intra}(h)$  в формуле (2) представляет собой среднее арифметическое от средних расстояний между всеми наблюдениями каждого класса и центроидом. Среднее расстояние между всеми наблюдениями в классе  $k$  и центроидом определяется как:

$$\text{Inner}(k) = \sum_{i=1}^{m_k} \frac{|X_i^k - r_0^k|}{m_k}. \quad (3)$$

<sup>1</sup> Статья написана на основе исследования: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3793810](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3793810).

Термин  $\text{Inner}(k)$  в формуле (3) указывает степени разделения данных внутри класса. Чем меньше индекс внутриклассовой компактности, тем меньше разница между данными внутри класса, содержащего  $h$  признаков. Кроме того, комбинация с  $h$  признаками обладает лучшей способностью различия между дефолтами и недефолтами.

### Индекс межклассового разделения

Индекс межклассового разделения определяется как:

$$\text{Inter}(h) = \frac{D(K)}{\beta(K)}, \quad (4)$$

где  $D(K)$  представляет собой среднее расстояние между центроидами, а  $\beta(K)$  представляет собой среднее расстояние от центроида каждого класса  $r_0^i$ :

$$D(K) = 2 \sum_{1 \leq i < j \leq K} \frac{|r_0^i - r_0^j|}{K(K-1)}, \quad (5)$$

$$\beta(K) = \sum_{1 \leq i \leq K} \frac{|r_0^i - r_0|}{K}. \quad (6)$$

Член  $\text{Inter}(h)$  в формуле (4) обозначает степени разделения между классами данных. Чем больше индекс межгруппового разделения, тем больше разница между классами. Комбинация с  $h$  признаками обладает лучшей способностью различия между дефолтами и недефолтами.

### Индекс достоверности

При учете как индекса внутриклассовой компактности, так и индекса межклассового разделения индекс достоверности можно определить как:

$$\text{VI}(h) = \exp \left\{ - \frac{\text{Inter}(h)}{\text{Intra}(h)} \right\}. \quad (7)$$

Индекс достоверности (VI) в формуле (7) показывает степени разделения между классами данных. Очевидно, что  $0 < \text{VI}(h) < 1$ ; меньший индекс достоверности указывает на то, что классы находятся дальше друг от друга. Оптимальная комбинация признаков может быть получена при минимальном значении индекса достоверности.

Процесс использования индекса достоверности для выбора оптимальной комбинации признаков выглядит следующим образом:

## Исследование кредитных рейтингов компаний КНР: качество и интерпретабельность

рассчитывается индекс достоверности для  $2^{m-1}$  комбинаций признаков с  $m$  признаками, затем размер всех индексов достоверности сравнивают и комбинация признаков, соответствующая наименьшему индексу достоверности, считается оптимальной.

Результаты (оптимальная комбинация из 26 признаков) показаны в табл. 1.

Таблица 1

### Оптимальная комбинация признаков

№	Признак	Вес
1	Соотношение активов и обязательств	0,0331
2	Коэффициент собственного капитала	0,0570
3	Счета к оплате	0,0304
4	Рентабельность собственного капитала	0,0429
5	Рентабельность инвестиционного капитала	0,0423
6	Операционные доходы/общая операционная выручка	0,0359
7	Налог на прибыль	0,0270
8	Операционный цикл	0,0372
9	Срок погашения кредиторской задолженности	0,0497
10	Срок погашения дебиторской задолженности	0,0289
11	Оборачиваемость оборотного капитала	0,0498
12	Количество уведомлений о задолженности	0,0279
13	Тип аудиторского заключения	0,0329
14	Показатель выплаты дивидендов	0,0346
15	Наличие изменений структуры собственного капитала	0,0347
16	Количество собраний наблюдательного совета	0,0663
17	ВВП	0,0436
18	Ежегодный подушный доход городских домохозяйств	0,0439
19	Общий объем гарантий/чистые активы	0,0625
20	Чистые активы на акцию	0,0397
21	Оборотный капитал/общие активы	0,0017
22	Материальные активы/общие активы	0,0255
23	Возраст председателя	0,0417
24	Эффективность внутреннего контроля	0,0503
25	Доступность аудиторских отчетов внутреннего контроля	0,0262
26	Вид нарушения	0,0342

## Построение оптимальной классификационной модели

Модель гравитации данных (DGM) сравнивает элемент выборки с «гравитацией» между различными классами выборки, чтобы определить тот класс, к которому он принадлежит.

Сила гравитации в рамках модели описывается следующим образом:

$$F = \frac{m_1 m_2}{r^2}, \quad (8)$$

где  $F$  — гравитация между двумя частицами данных;

$m_1$  — количество наблюдений в частице данных 1;

$m_2$  — количество наблюдений в частице данных 2;

частица данных — это тип единицы данных, состоящей из группы наблюдений.

Для каждой части данных клиента  $X_i$ ,  $k$  количество наблюдений с наименьшим евклидовым расстоянием выбирается из недефолтных и дефолтных наблюдений.

Гравитация между данными двух наблюдений обозначается как:

$$f = \frac{1}{r^2}. \quad (9)$$

Мы учитываем веса признаков при вычислении евклидова расстояния между двумя наблюдениями. Предполагается, что всего у нас  $h$  признаков и каждый признак — это  $x_i$  ( $i = 1, 2, 3, \dots, h$ ), а вес каждого признака равен  $w_i$  ( $i = 1, 2, 3, \dots, h$ ). Евклидово расстояние определяется как:

$$r = \sqrt{w_1^2(x_1^1 - x_1^2)^2 + w_2^2(x_2^1 - x_2^2)^2 + \dots + w_h^2(x_h^1 - x_h^2)^2}. \quad (10)$$

После этих вычислений для получения значений  $f$  между данными клиента  $X_i$  и данными дефолта и недефолта вычисляют общее значений  $2k$ .

Обозначим значения  $f$  между  $X_i$  и  $k$  недефолтных данных как  $f_0^1, f_0^2, f_0^3, \dots, f_0^k$ , а значения  $f$  между  $X_i$  и  $k$  дефолтных данных — через  $f_1^1, f_1^2, f_1^3, \dots, f_1^k$ . Добавляя значения  $f$ , гравитацию данных для  $X_i$  и  $S_0$  и  $X_i$  и  $S_1$  задаем как  $F_0$  и  $F_1$  соответственно, где:

$$F_0 = \sum_{i=1}^k f_0^i, \quad (11)$$

$$F_1 = \sum_{i=1}^k f_1^i. \quad (12)$$

## Исследование кредитных рейтингов компаний КНР: качество и интерпретабельность

Кредитный рейтинг  $X_i$  выражается как:

$$C_i^0 = \frac{F_0}{F_1 + F_0}. \quad (13)$$

Результаты модели гравитации данных (MDGM-1 и MDGM-2) в сравнении с традиционно используемыми в кредитных рисках моделями показывают лидерство применяемого подхода (табл. 2).

Таблица 2

### Ранги результатов модели

Serial number	Feature selection	Model	Rank					Mean	Final rank
			Accuracy	Type I error	Type I error	G-mean	AUC		
1	Yes	Logit	5,8	7,2	2,6	3,0	7,0	5,12	4
2		DT	7,2	6,0	7,6	7,8	3,4	6,40	9
3		LDA	6,4	6,8	5,8	5,6	5,8	6,08	8
4		$k$ NN	4,4	5,0	5,2	5,4	5,0	5,00	3
5		BPNN	4,4	3,2	6,2	6,2	6,0	5,20	5
6		SVM	5,6	4,8	5,0	5,0	7,8	5,64	7
7		ELM	6,0	6,0	6,0	6,2	3,6	5,56	6
8		MDGM-1	4,2	4,6	4,4	4,2	5,4	4,56	2
9		MDGM-2	1,0	1,4	2,2	1,6	1,0	1,44	1
10	No	Logit	4,4	5,6	4,0	4,6	7,6	5,24	5
11		DT	7,8	5,8	7,8	8,0	7,2	7,32	9
12		LDA	5,8	6,6	5,6	5,8	6,0	5,96	6
13		$k$ NN	3,2	2,2	5,0	4,0	4,8	3,84	3
14		BPNN	4,2	4,8	4,6	4,6	3,2	4,28	4
15		SVM	7,2	6,6	7,2	7,6	6,2	6,96	8
16		ELM	7,8	6,8	6,4	6,4	6,0	6,68	7
17		MDGM-1	3,6	5,2	2,8	2,8	3,0	3,48	2
18		MDGM-2	1,0	1,4	1,6	1,2	1,0	1,24	1

В прошлый раз мы говорили о том, на что обращать внимание при кредитовании крупных ИТ-компаний<sup>1</sup>. Публикация набрала много положительных отзывов, поэтому мы решили продолжить тему проектных рисков и специфических ИТ-рисков. Сегодня обсудим, как кредитовать ИП и небольшие компании, которые оказывают услуги в сфере ИТ или предлагают какой-то один продукт. Если клиент — ИП, банки, как правило, ограничиваются оценкой финансового положения и часто предлагают потребкредит. Но для таких клиентов можно создать отдельный продукт, сбалансированный по риску.

## Как оценить риски кредитования небольшого ИТ-бизнеса



Данил ДИНЦИС,  
менеджер проектов  
и портфелей, д.т.н.

### Кредитование ИП

— Данил, к нашей беседе мы подготовили выгрузку данных по ИП Санкт-Петербурга, заключивших госконтракты в сфере ИТ, и выделили несколько групп по видам деятельности. Группа, генерирующая больше всего денежных потоков, — это поставка оборудования, серверов, их подключение, обслуживание и т.п. Есть ли у нее специфические риски?

— Если говорить в целом об этой сфере, то, полагаю, риск заключается прежде всего в том, что это довольно узкое направление, которое существенно зависит от своего клиентского сегмента.

Что касается названной вами группы, то для нее могут быть характерны несколько рисков:

- риск задержки поставок оборудования;
- риск неработоспособности оборудования (связан с некачественной проработкой требований к оборудованию, к поставщику, к совместимости оборудования);
- тесно связанный с предыдущим риск нехватки компетенций.

— А как оценить компетенции заемщика? Возможно, обратить внимание на то, долго ли он работает на рынке?

<sup>1</sup> Как снизить риски кредитования ИТ-проектов в 2021 году // Риск-менеджмент в кредитной организации. 2021. № 1.

## Как оценить риски кредитования небольшого ИТ-бизнеса

— По моему мнению, это нерелевантный критерий. До открытия бизнеса заемщик мог работать в найме и получить там необходимые компетенции. Он мог привлечь в свой бизнес людей с высоким уровнем квалификации. Прямой корреляции между сроком присутствия на рынке и наличием опыта я не усматриваю.

**— Может быть, вместо пакета документов компании запрашивать трудовую книжку заемщика?**

— Запрашивать трудовую книжку, возможно, излишне, но посмотреть на персональный опыт самого ИП и его команды стоит — например, собрать отзывы, как это делают HR-специалисты. По сути, кредитовать следовало бы тех, кого вы бы наняли на работу.

Рекомендую обратить внимание на участие в профессиональных сообществах, партнерских ассоциациях и т.п.: активен ли заемщик в этих группах, пользуется ли он доверием? Если да, это поможет снизить риск недостатка компетенций или их потери (когда из маленькой компании уходит ключевой сотрудник).

**— Еще одна крупная группа — 1С-разработчики. Какие здесь могут быть нюансы?**

— Если для первой группы необходимость кредитования понятна (деньги нужны на закупку оборудования), то здесь целесообразно ставить вопрос: зачем такому ИП кредит? Если же цель обоснована, то подход к анализу заемщика может быть тем же — исследование индивидуального опыта и участия в профессиональных сообществах.

**— Есть ли способ «оцифровать» вес в профессиональном сообществе?**

— Да, для этого есть готовые метрики в инструментарии социальных графов. Например:

- 1) гомофиля (степень связей объекта с себе подобными);
- 2) сплоченность — степень связанности пользователей в сообществе;
- 3) центральность — степень влияния пользователя в группе (по числу связей);
- 4) взаимность («друг моего друга — мой друг»);
- 5) множественность (число множественных связей, например вуз и работа);
- 6) мост — объект, являющийся связью между кластерами (несколькими сообществами);

Принимая решение о кредитовании ИП, стоит собрать отзывы, как это делают HR-специалисты. По сути, кредитовать следовало бы тех, кого вы бы наняли на работу.

## Данил ДИНЦИС

7) расстояние — минимальное количество связей для того, чтобы установить взаимосвязь между двумя отдельными пользователями.

Эти метрики банк может включить в свои скоринговые модели, но важно настроить системное получение исходных данных и обеспечить их качество.

— **А какой может быть скоринговая модель для такого масштабного продукта?**

— Можно рассмотреть вариант на основе дерева решений. «Узлами», каждый из которых имеет свой вес, могут быть названные мной параметры: наличие индивидуального опыта, участие в профессиональных сообществах.

— **Есть ли специфические риски, связанные с продуктами 1С?**

— Эти продукты в своем профильном сегменте являются стандартами де-факто. 1С — обширнейшая экосистема, в которую входят и франчайзи, что уменьшает риски. В этой экосистеме ниже риски заказчика: он легко может поменять подрядчика. Но и риски подрядчика ниже — по той же причине взаимозаменяемости. Важно, что в экосистему 1С включена прозрачная многоуровневая система обучения и сертификации. Банк может проверить и наличие статуса партнера/франчайзи, и наличие сертификата. Это не является стопроцентной гарантией возврата кредита, но, полагаю, позволит снизить риски.

— **Учитывая взаимозаменяемость и легкость привлечения ресурсов, можно ли сказать, что количество сотрудников в данном случае не имеет значения? Допустим, ИП с тремя сотрудниками заключил контракт на 100 млн руб.**

— По моему мнению, стоп-фактором это не является.

— **Но это если технология известная и проработанная. А если нет? Например, ряд ИП занимаются разработкой и сопровождением малоизвестных специализированных систем.**

— В таком случае это может быть стоп-фактором. Риски сопряжены с тем, что потеря компанией одного из сотрудников часто означает потерю компетенции.

— **А насколько велик риск, что с появлением другой узкоспециализированной системы, причем в open source, деятельность компании закончится?**

Скоринговая модель для оценки риска ИП может быть построена на основе дерева решений. «Узлы», каждый из которых имеет свой вес, — это, например, такие параметры, как наличие индивидуального опыта, участие в профессиональных сообществах.

## Как оценить риски кредитования небольшого ИТ-бизнеса

— Надо понимать, что opensource — не бесплатный источник: он все равно требует развертывания, настройки, поддержки, развития. Это просто другая структура расходов (в первую очередь операционные расходы).

То есть речь идет об обычном рыночном риске, присущем деятельности любого заемщика.

### — Тогда каковы специфические риски таких ИП?

— Любые колебания на узкопрофильном рынке (новое регулирование, изменения в позиции профессионального сообщества и т.п.) быстро оказывают негативное влияние. Это обусловлено ограничением объемов микробизнеса и отсутствием сбалансированного портфеля.

**— Есть ли риск конкуренции со стороны крупных компаний? Например, если завтра такой же продукт появится у Microsoft или Google?**

— В ИТ риск конкуренции малого бизнеса с крупным относится к категории спекулятивных — это риски, у которых есть и позитивная, и негативная составляющие. Негативная более очевидна: работать с крупными подрядчиками часто дешевле, они более устойчивы и т.п. Но здесь стоит вспомнить наш предыдущий разговор. Большие компании объективно более «заформализованные» и не готовы подстраиваться под требования заказчика. Крупных заказчиков это часто не устраивает, и для них не так значима цена: их постоянные издержки могут быть сопоставимы со стоимостью тендера. Здесь микробизнес даже находится в выигрышной позиции по сравнению с крупным.

Многие заказчики любят работать поэтапно, «от достигнутого» (своего рода agile). С крупными подрядчиками они не всегда могут так организовать работу.

**— На форуме Scoring Day<sup>1</sup> говорилось о том, что банки не умеют кредитовать сельхозпроизводителей, особенно малых и средних. Как развить компетенцию по оценке таких заемщиков?**

— Полагаю, такие компетенции могут быть у онлайн-банков, ориентированных на работу с МСБ. Здесь соединяются два факто-ра — целевая аудитория и мощная ИТ-экспертиза.

В ИТ риск конкуренции малого бизнеса с крупным относится к категории спекулятивных — это риски, у которых есть и позитивная, и негативная составляющие. Большие компании не готовы подстраиваться под требования заказчика. Здесь микробизнес даже находится в выигрышной позиции.

<sup>1</sup> Scoring Day 2021 Весна. 9-я межотраслевая конференция (scorconf.ru).

## Данил ДИНЦИС

— Таким образом, в банке должны быть Product Owners, каждый из которых знает свой сегмент?

— Да, важна специализация: «железо», основные системы, Service Desk и др. Организовать такую работу поможет ITIL (IT Infrastructure Library). Стоит обратить внимание, есть ли у микробизнеса базовые элементы IT Service Management, которые определяют работу с клиентами, прохождение заявок и прочие основные этапы. Это делает более предсказуемой работу заемщика не только с заказчиком, но и с банком.

### Кредитование небольших быстроразвивающихся компаний

— В обзоре СПАРК-Интерфакс<sup>1</sup> говорится о том, что для отнесения ИТ-компаний к динамично развивающимся и финансово устойчивым она должна отвечать нескольким критериям. Во-первых, среднегодовой прирост выручки должен превышать 10% в постоянных ценах в течение последних трех лет. Во-вторых, выручка должна быть более 70 млн руб. — это уровень, с которого финансовая устойчивость такой компании существенно возрастает. В-третьих, чистые активы компании в последний отчетный год не должны быть отрицательными. И наконец, у нее должен быть средний или низкий индекс финансового риска<sup>2</sup>. Пока, по данным СПАРК-Интерфакс, этим критериям соответствует около 3% ИТ-компаний. Как найти «бриллианты»?

— Группой партнеров российского Microsoft разработаны Рекомендации по выбору подрядчика для поддержки ИТ-инфраструктуры бизнеса. Эти рекомендации могут использовать и банки для оценки потенциального заемщика — узкопрофильной или многопрофильной компании.

В обзоре СПАРК-Интерфакс есть компании, которые предоставляют, например, веб-доступ или услуги SMS-рассылки. Если компания специализируется в узкой области, она накапливает глубокую экспертизу и соответственно более надежна. У нее отлажены процессы, она может проводить более гибкую ценовую политику за счет большего числа клиентов.

Слабые стороны такой компании в том, что она зависит от узкого сегмента рынка и мгновенно реагирует на любые изменения

Группой партнеров российского Microsoft разработаны Рекомендации по выбору подрядчика для поддержки ИТ-инфраструктуры бизнеса. Эти рекомендации могут использовать и банки для оценки потенциального заемщика — узкопрофильной или многопрофильной компаний.

<sup>1</sup> Сектор ИКТ в России: начало ускорения? СПАРК, 2021.

<sup>2</sup> Об индексе финансового риска см., например: Мунерман И., Воронцов И. Как прогнозировать неплатежеспособность заемщика в условиях недостаточных или недостоверных данных? // Банковское кредитование. 2019. № 2.

## Как оценить риски кредитования небольшого ИТ-бизнеса

на рынке. Могу утверждать это, в том числе основываясь на своем опыте работы в компании, которая специализировалась на услугах для страхового рынка. Были периоды, когда мы были вынуждены практически в два раза увеличивать штат, чтобы хоть как-то справиться. Это, кстати, тоже большой риск, потому что при резком увеличении штата теряется управляемость (о чем говорят и «модные» практики DevOps, Scrum и пр.). Когда же на рынке «затишье» — возникают проблемы с cash flow.

К слабым сторонам узкоспециализированных компаний стоит отнести и ограниченный кругозор. В Рекомендациях отмечается низкое качество менеджмента в таких компаниях, но я бы не утверждал это столь категорично. По крайней мере, в небольших компаниях, с которыми я работаю последние 7–8 лет в качестве консультанта, это не так.

### От редакции

Даже в «провальном» II квартале 2020 г. компании сектора ИКТ (информационно-коммуникационных технологий) смогли нарастить оборот почти на 4%, притом что общероссийский показатель упал на 11%. Во второй половине 2020 г. рост оборота сектора ускорился до 10%.

По данным СПАРК-Интерфакс, в 2020 г. совокупная выручка компаний сектора увеличилась в среднем на 16% — до 5655 млрд руб., чистая прибыль на 49% — до 661 млрд руб. Рентабельность сектора по чистой прибыли выросла в среднем с 9,2 до 11,7%. Показатель «чистый долг/ЕВИТ» снизился с 2,3x до 1,8x.

Доля компаний с индексом платежной дисциплины «Высокий риск» в конце 2020 г. была менее 6%, средний индекс «Низкий риск» составил 82%. Количество конкурсных производств на 1000 компаний оказалось почти в 3 раза ниже среднего российского уровня.

Из рис. 1 видно, что в 2017–2020 гг. среднегодовые темпы прироста выручки, прибыли от продаж и чистой прибыли динамично развивающихся компаний ИКТ по сравнению с данными 2016–2019 гг. выросли, но уровня 2015–2018 гг. не достигли. Как показывает табл. 1, улучшение показателей роста чистой прибыли обеспечил кластер телекоммуникаций. Однако наиболее высокий уровень прибыльности отмечался в сфере информационных технологий — 19,5% по сравнению с 14,5% в прошлом периоде.

Средняя рентабельность чистой прибыли динамично развивающихся компаний ИКТ достигла 18,2% по сравнению с 13,3% в 2016–2019 гг. и 3,5% в 2015–2018 гг. Данные о рентабельности по сегментам приведены в табл. 2.

## Данил ДИНЦИС

Рисунок 1

### Средние темпы роста выручки и прибыли динамично развивающихся компаний ИКТ (в постоянных ценах)

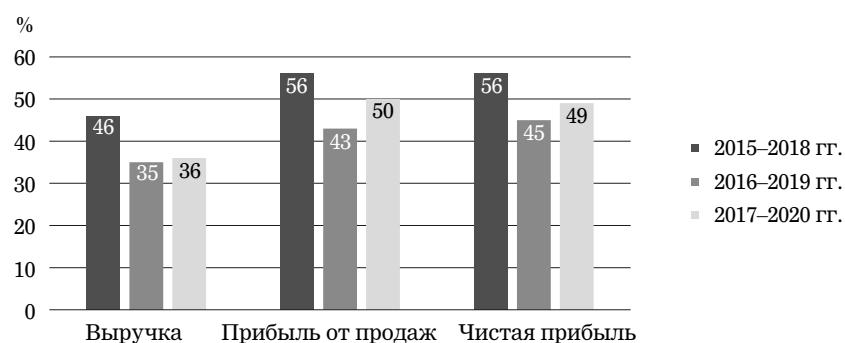


Таблица 1

### Средние темпы роста выручки и прибыли динамично развивающихся компаний ИКТ по сегментам (в постоянных ценах)

Отрасль	Телекоммуникации		Информационные технологии	
	2016–2019 гг.	2017–2020 гг.	2016–2019 гг.	2017–2020 гг.
Выручка	43%	35%	42%	36%
Чистая прибыль	75%	82%	53%	49%

Источник: СПАРК-Интерфакс.

Таблица 2

### Рентабельность чистой прибыли динамично развивающихся компаний ИКТ на конец периода

	2015–2018 гг.	2016–2019 гг.	2017–2020 гг.
Телекоммуникации	7,1%	11,3%	10,4%
Информационные технологии	14,1%	14,5%	19,5%

Источник: СПАРК-Интерфакс.

При отнесении компаний к сектору ИКТ СПАРК-Интерфакс руководствуется основным ОКВЭД: это должен быть ОКВЭД 61 (деятельность в сфере телекоммуникаций), 62 (разработка компьютерного

## Как оценить риски кредитования небольшого ИТ-бизнеса

ПО, консультационные услуги в данной области и другие сопутствующие услуги) или 63 (деятельность в области информационных технологий).

По оценке СПАРК-Интерфакс, устойчивый рост динамично развивающихся компаний ИКТ обеспечен в основном собственными ресурсами, а не заемными средствами. Эти компании в 2019–2020 гг. не имели совокупного чистого долга, то есть размер высоколиквидных активов превышал задолженность по займам и кредитам. Чистый долг был только у телекоммуникационных компаний. Их средний показатель «чистый долг/ЕВИТ» составлял 2,4x — это значительно ниже показателя всего телекоммуникационного сектора (4,8x).

**— А сколько человек должно быть в такой идеальной команде с большим потенциалом? И кем должен быть владелец: «технарем» или управленцем?**

— Не думаю, что на первый вопрос есть однозначный ответ. Продукты бывают разные, масштаб компании пропорционален масштабу продуктов. Принцип, в соответствии с которым в команде должно быть 7 плюс-минус 2 человека, справедлив для команды разработки, но для компаний в целом неприменим.

Нужна профессиональная оценка того, насколько состав — и количественный, и квалификационный — позволяет обеспечить нужные компетенции исходя из сути продукта, из того количества технологических, функциональных областей или бизнес-областей, которые он охватывает.

Что касается второго вопроса, то стоит уточнить: имеется в виду руководитель компании или руководитель команды?

**— Очевидно, руководитель команды должен сам «кодить».**

— Ни в коем случае. Если у него есть технический бэкграунд, то есть он может консолидировать мнения, разговаривать с командой на одном языке, — это серьезный плюс. А вот если он продолжает «кодить» сейчас, я бы это рассматривал как очень существенную угрозу.

**— То есть он не на то будет тратить время?**

— Да, абсолютно верно. И его техническое мнение как руководителя может превалировать, что приводит к очень нехорошим последствиям. Оптимальный вариант — если он когда-то работал в профильной области и удерживает уровень достаточной технической компетенции.

## Данил ДИНЦИС

### От редакции

Рекомендации группы партнеров российского Microsoft не применимы напрямую к ИТ-услугам проектного характера, таким как внедрение ERP, CRM или другой корпоративной системы, разработка заказного специального ПО и т.п. Описываемая модель хорошо работает для таких услуг, как комплексное сопровождение ИТ-инфраструктуры в целом или ее составных частей (оборудования, системного и прикладного ПО).

Риски узкопрофильных и многопрофильных ИТ-компаний можно оценить при помощи табл. 3 и 4.

Таблица 3

### SWOT-анализ ИТ-компаний, специализирующихся на узкой области ИТ-услуг

Сильные стороны	Слабые стороны
<ol style="list-style-type: none"> <li>Высокий уровень профессионализма в области специализации.</li> <li>Возможно предоставление относительно низких цен на услуги за счет отлаженных процессов.</li> <li>Возможно предоставление относительно низких цен на оборудование и ПО за счет особых условий работы с поставщиками</li> </ol>	Ограниченный кругозор, невозможность комплексного подхода к поддержке ИТ-инфраструктуры
<b>Возможности повышения эффективности работы</b>	<b>Угрозы эффективности работы</b>
Целесообразно использовать для поддержки тех частей ИТ-инфраструктуры, для которых содержание собственного штата специалистов менее эффективно	Низкое качество менеджмента

Таблица 4

### SWOT-анализ многопрофильных ИТ-компаний

Сильные стороны	Слабые стороны
<ol style="list-style-type: none"> <li>Высокий уровень профессионализма в областях специализации.</li> <li>Возможно предоставление относительно низких цен на услуги за счет отлаженных процессов.</li> <li>Возможно предоставление относительно низких цен на оборудование и ПО за счет особых условий работы с поставщиками.</li> <li>Сокращение транзакционных издержек («принцип одного окна»)</li> </ol>	(Возможно) Относительно более высокая стоимость услуг за счет более высоких затрат на управление компанией
<b>Возможности повышения эффективности работы</b>	<b>Угрозы эффективности работы</b>

## Как оценить риски кредитования небольшого ИТ-бизнеса

Окончание табл. 4

Возможности повышения эффективности работы	Угрозы эффективности работы
Комплексное предоставление услуг избавляет от возможных конфликтов при возникновении инцидентов «на стыке» разных элементов ИТ-инфраструктуры (например, если непонятно, чем вызвана проблема — оборудованием, системным или прикладным программным обеспечением)	Трудно определить обоснованность цены за комплексные услуги

А на рис. 2 представлена любопытная квалификация типов подрядчиков. Требования банка к предоставляемой потенциальными заемщиками информации можно формулировать так, чтобы выявлять принадлежность к тому или иному типу и отсекать относящихся к нежелательным типам.

Рисунок 2

### ИТ-компании на графике «цена-качество»



#### Примечания:

Основная масса исполнителей попадет в категорию **середняков** – подрядчиков, обеспечивающих среднее качество при среднем уровне цен.

Ярко выделяется на общем фоне **звезда** – подрядчик, оказывающий услуги заметно более высокого качества по высоким ценам.

Но еще более ярко будет выделяться **трудоголик** – подрядчик, при заметно более высоком качестве предлагающий средние или незначительно более высокие цены. Таких исполнителей любят, но они остаются такими только до тех пор, пока количество их заказчиков не позволит им «возвеститься» к звездам.

**Начинающий** – подрядчик, при заметно более низких ценах обеспечивающий среднее или незначительно более низкое качество.

**Селебрити** – подрядчик, при заметно более высоких ценах обеспечивающий качество среднее или незначительно выше среднего. Распространенная, хотя и нелюбимая заказчиками из МСБ категория.

Редкая категория – **лузер**: подрядчик, предлагающий низкие цены, но не способный поддерживать даже среднее качество. Лозунг: «Мы делаем вид, что работаем, а вы – что нам платите». Наконец, печальная судьба исполнителей типа **«спена»**: подрядчик, предлагающий средние цены, но не способный поддерживать даже среднее качество. Такие исполнители обслуживают клиента, пока тому не подвернется хотя бы сердняк.

## Данил ДИНЦИС

Преимущества начинающих — в высокой заинтересованности и увлеченности, а также в том, что им надо показывать высокое качество при сравнительно низких ценах. С точки зрения банка это маркер.

Дальше банковские андеррайтеры должны решать, чего больше — рисков или потенциальных выгод.

- Рекомендуется проверять, в частности:
- добросовестность выставления цены — при помощи прове-рочных звонков. Вначале путем запроса информации собираются данные от различных исполнителей для определения средних рыночных цен, затем путем запроса информации о возможности получения рекомендаций собираются контактные данные, а по ним — информация о добросовестности ценообразования;
  - полноту описания методологии. Качество выше среднего возможно обеспечить только при высоком уровне организации процессов, показателем которого является их формализация (опи-сание);
  - адаптивность — путем референс-визитов. Рекомендуется посетить действующего клиента и в ходе переговоров выяснить на конкретных примерах порядок и скорость реакции на негатив;
  - срок работы с клиентами. Длительный срок работы с одним клиентом легко проверяется и является показателем удовлетво-ренности ценой и качеством;
  - наличие уточняющих вопросов, что служит показателем опыта.
- Вся эта информация легко верифицируется, а соответствующие названным критериям подрядчики с высокой вероятностью будут относиться к типам «трудоголик», «середняк» или «начинающий».

**— Судя по графику «цена-качество», банки должны стараться находить трудоголиков.**

— Если говорить о сегменте небольших профильных компаний, то да. Но я бы не сбрасывал со счетов начинающих, особенно сейчас. Культура стартапов, в том числе внутренних, развивается — возьмем для примера те же Яндекс, МТС, Сбер, которые заявляют о наличии внутренних стартапов.

Преимущества начинающих, во-первых, в их высокой заинтересованности и увлеченности, а во-вторых, в том, что им надо показывать высокое качество при сравнительно низких ценах. Для заказчика это очень неплохо. С точки зрения банка, по крайней мере, это маркер. Дальше банковские андеррайтеры должны решать, чего больше — рисков или потенциальных выгод.

**— То есть задача банка — найти людей с высокой мотивацией?**

— Да, и притом найти их проще, потому что они стараются активно заявлять о себе в сети и на профильных мероприятиях. В прошлый раз мы говорили о репутационном факторе — для начинающих этот

## Как оценить риски кредитования небольшого ИТ-бизнеса

фактор особенно значим. Речь идет о личной репутации. Понятно, что компанию на рынке не знают и вряд ли будут запоминать, пока основатель не станет звездой. Более того, сегодня он может открыть одну компанию, а завтра другую. Но можно ли иметь дело с этим конкретным человеком и этой конкретной командой — вот что важно.

### От редакции

Согласно результатам ежегодного исследования «СПАРК: Быстро-растущие компании (2016–2020 гг.)»<sup>1</sup>, количество «газелей» из сектора ИКТ составило 42 компании, в 2015–2019 гг. их было всего 20. На разработку ПО пришлось 83% от общего числа «газелей», на телекоммуникации — 17%.

Среди быстрорастущих компаний — WaveAccess, СимбирСофт (разработчики ПО), Syntellect (поставщик решений для ЭДО), МаксимаТелеком (оператор крупнейшей в Европе публичной сети Wi-Fi). Стоит учесть, что к «газелям» сектора ИКТ СПАРК-Интерфакс относит и такие компании, как ЛитРес, PickPoint и Ivi.

— Получается, что банкам надо находить либо узкоспециализированные компании, либо компании, за которыми стоит «понятный» человек, и уже на него делать ставку. Потому что сейчас конкуренция на банковском рынке очень большая. И если у вас выручка 3–5 миллиардов в год, вы завалены банковскими предложениями. Новый банк в этот сегмент вряд ли сможет зайти.

— Что еще важно: однозначно стоит обращать внимание на наличие отзывов, особенно публикуемых на сайте. Распространенная практика, когда, выполнив работу, подрядчик просит отзыв на бланке компании. Наличие таких отзывов явно повышает доверие к подрядчику, а значит, повышает его устойчивость.

Как и в случае с заемщиками-ИП, обратите внимание, входит ли компания в профильные ассоциации, партнерства. Очевидно, что она не сможет покрыть все аспекты, которые нужны заказчику, поэтому ей придется интегрироваться с другими компаниями. Оцените, насколько она готова к такой интеграции.

Также важна готовность к референс-визитам. Полагаю, представители банка тоже могли бы совершить референс-визит: это не требует больших вложений и занимает два-три часа.

Оцените готовность потенциального заемщика к референс-визитам. Представители банка тоже могли бы совершить референс-визит: это не требует больших вложений и занимает два-три часа.

<sup>1</sup> См. также: Мунерман И., Волосов И. «Газели» накануне вымирания? Быстрорастущая компания как потенциальный заемщик // Банковское кредитование. 2020. № 5; «Газели» как потенциальные заемщики: что изменилось за пандемийный год // Банковское кредитование. 2021. № 4.

## Данил ДИНЦИС

В моей практике был случай: мы участвовали в тендере на страховом рынке. У нас запросили референс-визит. Я обратился к ИТ-директору компании, с которой мы реализовали уже несколько проектов, и он дал согласие принять своих коллег из другой компании, но с характерной оговоркой: «Мы не ваши рекламные агенты и расскажем про вас все как есть». Мы выиграли тендер, и когда у нас установились рабочие отношения с новой компанией, я узнал, что мы были одни из немногих или даже единственных, кто предоставил референс-визит. Причем наших новых контрагентов это даже насторожило, и они решили, что если нас будут расхваливать, то это заведомый обман. Однако им назвали как сильные, так и слабые наши стороны, предупредили, где нас нужно контролировать. Это не только стало доказательством реальности нашей практики на рынке, но и помогло контрагентам понять, на что обратить внимание при работе с нами.

Применительно к банковской практике, если контрагент действительно выступает в роли «рекламного агента» и называет только положительные стороны, это можно считать нехорошим признаком.

**— А если компания начинающая, в ней работают талантливые мотивированные люди, но они плохо зарабатывают деньги? Как понять, есть ли у нее потенциал, и спрогнозировать платежеспособность?**

— Здесь желательно смотреть уже не на техническую команду, а на команду финансовую и маркетинговую. Я сотрудничаю с акселератором стартапов, который при мне выпустил две команды. И в обоих случаях в ходе акселерационного процесса команда брала в соучредители кого-то из сферы маркетинга. Например, студенческая команда подбирала опытного маркетолога более солидного возраста. То есть product-driven подхода недостаточно: речь о бизнесе, а не о чистой разработке. 

**Интервью провел Владимир КОЗЛОВ**

По итогам нашего анализа hh.ru, в требованиях работодателя SQL занимает место выше Python. А учитывая, что в риск-менеджменте становится все больше данных и методы работы с ними должны знать уже не только рисковики, но и экономисты и кредитные аналитики, — мы, не претендуя на изложение продвинутых подходов, расскажем о том, что пока известно не всем и в чем некоторые допускают ошибки.

## SQL-Join: разбираемся один раз и надолго

Возьмем две таблицы — список клиентов банка и список их текущей задолженности на определенную дату.

Создадим базу:

```
CREATE TABLE clients (
    id INT(11) NOT NULL AUTO_INCREMENT,
    first_name VARCHAR(50) NOT NULL COLLATE 'utf8_unicode_ci',
    last_name VARCHAR(50) NOT NULL COLLATE 'utf8_unicode_ci',
    email VARCHAR(100) NOT NULL COLLATE 'utf8_unicode_ci',
    birthdate DATE NOT NULL,
    added TIMESTAMP NOT NULL DEFAULT CURRENT_TIMESTAMP,
    PRIMARY KEY (id),
    UNIQUE INDEX email (email)
)
COLLATE='utf8_unicode_ci';
CREATE TABLE balance (
    id INT(11) NOT NULL AUTO_INCREMENT,
    client_id INT(11),
    balance NUMERIC NOT NULL,
    date DATE NOT NULL,
    PRIMARY KEY (id)
)
COLLATE='utf8_unicode_ci';
```



**Владимир КОЗЛОВ,**  
FRM, консультант  
по риск-менеджменту

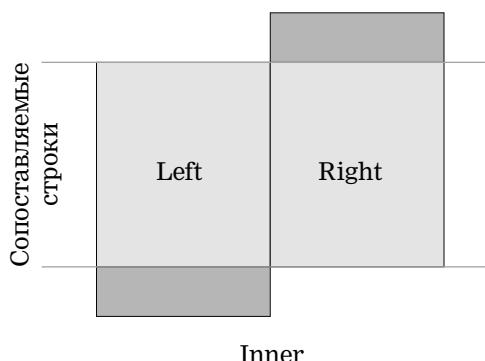
Как это часто бывает в банках, на `client_id INT(11)` нет ограничения `NOT NULL`, поэтому иногда попадаются клиенты из других баз или таблиц либо вовсе не идентифицированные лица. В `balance` у нас 10 клиентов без идентификаторов клиента, а в `clients` — 10 клиентов без информации о балансе.

## Владимир Козлов

### Базовый уровень

Чтобы разобраться с этим один раз и надолго, проследим исполнение наших скриптов в Microsoft Excel.

Для того чтобы получить только тех клиентов, которые есть в обеих таблицах, используем Inner join:



Скрипт:

```
SELECT
    b.id,
    b.client_id,
    b.balance,
    b.date,
    c.first_name,
    c.last_name
FROM
    balance b
INNER JOIN
    clients c
    on c.id = b.client_id;
```

Примечательно, что это эквивалент кода:

```
SELECT
    b.id,
    b.client_id,
    b.balance,
    b.date,
    c.first_name,
    c.last_name
FROM
    balance b, clients c
where c.id = b.client_id;
```

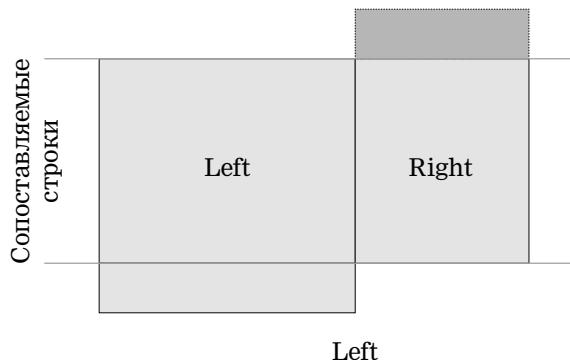
## SQL-Join: разбираемся один раз и надолго

Получилось 90 строк вида:

<b>id</b>	<b>client_id</b>	<b>balance</b>	<b>date</b>	<b>first_name</b>	<b>last_name</b>
10	10	0,00	2021-01-05	Holly	Hoppe
11	11	9 854 627,00	2021-01-20	Ezequiel	Leannon
12	12	251 784,00	2021-01-25	Frankie	Kris
13	13	39,00	2021-01-15	Natalie	Gulgowski
14	14	5 602,00	2021-01-16	Javon	Luettgen
15	15	0,00	2021-01-22	Mac	Breitenberg
16	16	69 400,00	2021-01-12	Norberto	Conner
17	17	558,00	2021-01-27	Giuseppe	Bode

Все строки заполнены. В этой выборке нет клиентов без баланса и нет балансовых записей без идентификатора клиента.

Для того чтобы «прицепить» задолженность ко всем клиентам из таблицы клиентов, используем Left join:



Left

Это наиболее часто используемый запрос. Скрипт:

```

SELECT
    b.id,
    b.client_id,
    b.balance,
    b.date,
    c.first_name,
    c.last_name
FROM
    clients c
LEFT JOIN
    balance b
on c.id = b.client_id;

```

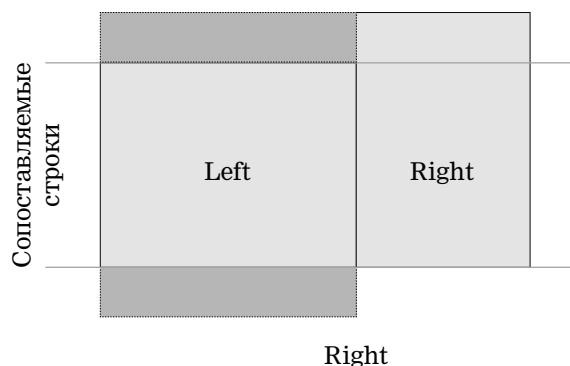
## Владимир Козлов

Мы получили 100 строк. Примечательна нижняя часть полученной таблицы:

<b>id</b>	<b>client_id</b>	<b>balance</b>	<b>date</b>	<b>first_name</b>	<b>last_name</b>
95	95	938 332,00	2021-01-05	Nikita	Wuckert
97	97	82 677,00	2021-01-03	Camden	Mosciski
98	98	689,00	2021-01-25	Aimee	Parker
99	99	4 141 120,00	2021-01-12	Flo	Price
100	100	365 504,00	2021-01-12	Laisha	Keefe
				Elliot	Roob
				Bria	Schuppe
				Gideon	Windler
				Emily	Kessler
				Heidi	Satterfield
				Roslyn	Eichmann
				Anabelle	Keefe
				Brenda	Kiehn

На 90-м клиенте полное соответствие таблиц закончилось. Так как основная таблица у нас в запросе клиентская — мы получили клиентов без баланса (наверное, в бизнес-смысле это клиенты с балансом 0).

Сравним с тем, что такое Right Join:



Скрипт:

```
SELECT
    b.id,
    b.client_id,
    b.balance,
    b.date,
    c.first_name,
    c.last_name
FROM
    clients c
```

## SQL-Join: разбираемся один раз и надолго

RIGHT JOIN

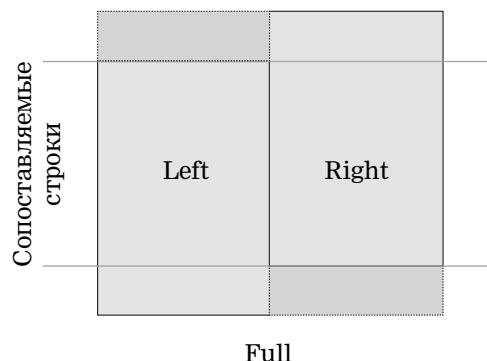
```
balance b
on c.id = b.client_id;
```

Опять смотрим на концовку таблицы. Получаем тех самых неидентифицированных клиентов:

<b>id</b>	<b>client_id</b>	<b>balance</b>	<b>date</b>	<b>first_name</b>	<b>last_name</b>
97	95	938 332,00	2021-01-05	Nikita	Wuckert
98	97	82 677,00	2021-01-03	Camden	Mosciski
99	98	689,00	2021-01-25	Aimee	Parker
100	99	4 141 120,00	2021-01-12	Flo	Price
101		100,00	2021-01-11		
102		2 000,00	2021-01-11		
103		4 860,00	2021-01-11		
104		120,00	2021-01-11		
105		546 620,00	2021-01-11		
106		0,00	2021-01-11		
107		0,00	2021-01-11		
108		78 540,00	2021-01-11		
109		0,00	2021-01-11		
110		0,00	2021-01-11		

В нашей dummy-базе им выдали средств более чем на 600 тыс. руб. На практике иногда счет идет на миллиарды — впрочем, данные хранятся в каких-то других базах, о которых знают ваши предшественники-рископики и местный DBA.

Наконец, чтобы срастить обе таблицы вне зависимости от их пересечений, используем Full join:



## Владимир Козлов

Скрипт:

```
SELECT
    b.id,
    b.client_id,
    b.balance,
    b.date,
    c.first_name,
    c.last_name
FROM
    balance b
FULL JOIN
    clients c
    on c.id = b.client_id;
```

Результат:

<b>id</b>	<b>client_id</b>	<b>balance</b>	<b>date</b>	<b>first_name</b>	<b>last_name</b>
97	97	82 677,00	2021-01-03	Camden	Mosciski
98	98	689,00	2021-01-25	Aimee	Parker
99	99	4 141 120,00	2021-01-12	Flo	Price
102		2 000,00	2021-01-11		
103		4 860,00	2021-01-11		
104		120,00	2021-01-11		
105		546 620,00	2021-01-11		
106		0,00	2021-01-11		
107		0,00	2021-01-11		
108		78 540,00	2021-01-11		
109		0,00	2021-01-11		
110		0,00	2021-01-11		
				Elliot	Roob
				Bria	Schuppe
				Gideon	Windler
				Emily	Kessler
				Heidi	Satterfield
				Roslyn	Eichmann
				Anabelle	Keefe
				Brenda	Kiehn
				Myrtis	Eichmann

## SQL-Join: разбираемся один раз и надолго

Ровно такой же результат можно достичнуть сочетанием left join и union all:

```
SELECT
    b.id,
    b.client_id,
    b.balance,
    b.date,
    c.first_name,
    c.last_name
FROM
    clients c
LEFT JOIN
    balance b
on c.id = b.client_id
union all
select
    b.id,
    b.client_id,
    b.balance,
    b.date,
    null,
    null
from balance b
where b.client_id is null
```

Этот скрипт не используют для извлечения данных, то есть в плане построения модели он бесполезен. Однако его можно и нужно использовать, когда необходимо найти ошибки в базе — а именно пустые значения ключей в связанных таблицах.

Особенно хорошо он помогает, когда имеется более 3–4 таблиц, которые необходимо объединить. Несмотря на то что результат может получится крайне массивным, порой остается только удивляться, откуда в нашей считавшейся «надежной» базе столько дубликатов или нераспознанных значений.

### Проверочное задание

Ответьте на вопрос: сколько строк выдаст запрос и почему?

```
SELECT
    b.id,
    b.client_id,
    b.balance,
    b.date,
    c.first_name,
    c.last_name
FROM
    balance b, clients c
```

## Владимир КОЗЛОВ

Ответ: 10 000, так как запрос построит Join по каждой строке обоих таблиц, что хорошо видно при explain-анализе запроса:

	Operation	Params	Rows
[]	↳ Select		
	↳ Nested Loops (Nest)	10000	
	[] Full Scan (Seq Sc table: balance;)	100	
	↳ Temporary (Mat)	100	
	[] Full Scan (Seq table: clients;)	100	

Подведем итог. Мы используем left join, когда надо обогатить данные основной таблицы данными иной таблицы. Проблемный момент left join — неидентифицированные данные дополнительной таблицы, они будут проигнорированы. Right join — обратная операция. Inner join — поиск полного пересечения по указанным ключам двух таблиц и при данных с невысоким качеством подготовки выдаст меньше всего строк. Full join используется или для поиска неучтенных значений в базе, или когда качество данных очень высокое (ошибки в базе минимальны, имеются ограничения на значения и т.п.). Union all используется для добавления строк, а не столбцов, в результат запроса, при этом по количеству столбцов он должен совпадать с основным запросом. 

Методический журнал

## Банковское кредитование

### Читайте в № 6/2021:

Роман БОЖЬЕВ, Никита ФЕДОТОВ, Александр ПИСКОТИН, *Объединенное Кредитное Бюро*

#### **Риск-факторы как фактор неэффективности кредитного конвейера**

Рисковые процедуры обладают большим потенциалом для улучшения ключевых метрик (уровня одобрения, времени принятия решения и пр.). Но есть основания предполагать наличие систематических ошибок внутри таких процедур. Наша задача — найти и исследовать эти ошибки.

Антон ВОВК, Иван ГУЛЕНКОВ, *Общественный совет при Росреестре*

#### **Как банкам снизить риск приостановки регистрации ипотеки**

При некорректном описании объекта недвижимости в договоре ипотеки банк получит уведомление о приостановке регистрации или даже об отказе в регистрации. На какие параметры объекта обратить внимание, чтобы этого избежать? Какие новые инструменты позволят банкам системно решить проблемы приостановок?

Леонид БЕЛЬЧЕНКО, Александра ВЕРОЛАЙНЕН, «Эксперт РА»

#### **Секьюритизация — все еще не использованный резерв для развития кредитования**

Объем рынка секьюритизации на конец III квартала 2021 г. достиг 600 млрд руб., но распределен крайне неравномерно между ипотекой и другими активами. Ликвидация регуляторных диспропорций поможет банкам привлечь средства для кредитования МСБ, проектного финансирования и поддержки экономически значимых сегментов.

Методический журнал

## Внутренний контроль в кредитной организации

### Читайте в № 4/2021:

Елена РОЗАНОВА, *ООО «РИСКФИН»*

#### **Положение № 716-П: предъявите ваши документы**

Положение № 716-П и разъяснения к нему установили «бенчмарк» требований на уровне крупнейших банков.

Анализ изменений показывает, что «пропорциональное регулирование» оказалось не таким уж пропорциональным. Как за прошедший год поменялась ситуация и как изменился «корпус текстов», связанных с требованиями к СУОР?

Михаил КАРАТАЕВ, банковский эксперт

#### **Платформа «Знай своего клиента»: как найти баланс между срочным и важным**

Банки стоят перед выбором: срочно исполнять изменившиеся нормы обязательной отчетности или сосредоточиться на управлении рисками сомнительных сделок. Цель статьи — помочь банкам понять скрытые особенности методологии нововведений для эффективного распределения ресурсов и выбора приоритетов.

Снежана ГАЗИЯН, *CIA, AIRC*

#### **Контрольные самооценки как инструмент непрерывного внутреннего контроля (на примере ПОД/ФТ/ФРОМУ)**

Все больший вес приобретают консультационные функции системы внутреннего контроля, координация и поддержка изменений. Инструментом для решения новых задач является опора на контрольные самооценки. В статье приведен практический пример подхода к формированию анкеты для такой самооценки.

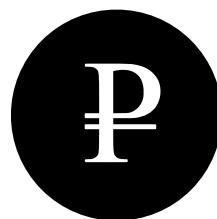
# Готовые решения ваших задач – в системе «РегламентБанк»!



«РегламентБанк» –  
база профессиональных изданий  
и методических пособий  
для банковских специалистов



Готовый опыт ваших  
коллег и ведущих  
банковских  
экспертов



Рекомендации  
аудиторов  
и консультантов



Комментарии  
и разъяснения  
от представителей  
Банка России



Четкие руководства  
к действию  
на основе анализа  
новых нормативных  
документов

Издательский дом «Регламент»  
более 15 лет занимается  
методическим сопровождением  
работы банковских специалистов



Все журналы и пособия ИД «Регламент»  
можно читать через систему «РегламентБанк»