



# ТЕХНИЧЕСКИЕ

# СПРАВКИ И РУКОВОДСТВА

## Налоговое администрирование

### Основные возможности аналитики для управления комплаенс-рисками

Джошуа Аслетт, Густаво Гонсалес, Стюарт Хэмилтон и Мигель Печо



---

## ТЕХНИЧЕСКИЕ СПРАВКИ И РУКОВОДСТВА

# Налоговое администрирование Основные возможности аналитики для управления комплаенс-рисками

Джошуа Аслетт, Густаво Гонсалес, Стюарт Хэмилтон и Мигель Печо  
Разрешено к распространению Абдельхаком Сенхаджи

### **В настоящей технической справке и руководстве рассматриваются следующие вопросы:**

- Что понимается под аналитикой и для чего она используется в налоговом администрировании?
- Как аналитика поддерживает процессы управления комплаенс-рисками (УКР)?
- Почему статистика и наука о данных имеют важное значение для современной аналитики?
- Какие данные, технологии и инструменты обычно используются для поддержания аналитических возможностей в области УКР?
- Какие аналитические функции, данные и знания необходимы для УКР?

© 2024 International Monetary Fund  
Перевод на русский язык ©2024 Международный валютный фонд  
Издание на русском языке подготовлено Службой переводов МВФ  
Дизайн обложки: Отдел креативных решений, ДКО МВФ

**Cataloging-in-Publication Data**  
**IMF Library**

Names: Aslett, Joshua, author. | González Amilivia, Victor Gustavo, author. | Hamilton, Stuart Gordon, author. | Pecho, Miguel, author. | International Monetary Fund, publisher.

Title: Tax administration : essential analytics for compliance risk management / Joshua Aslett, Gustavo González, Stuart Hamilton, and Miguel Pecho.

Other titles: Essential analytics for compliance risk management. | Technical notes and manuals.

Description: Washington, DC : International Monetary Fund, 2024. | Feb. 2024. | TNM/2024/01. | Includes bibliographical references.

Identifiers: ISBN:

9798400260063	(paper)
9798400260124	(ePub)
9798400256264	(Web PDF)
9798400272240	(Russian, paper)

Subjects: LCSH: Tax administration and procedure. | Public administration and public policy.

Classification: LCC HJ2305.A8 2024

**ЗАЯВЛЕНИЕ ОБ ОТКАЗЕ ОТ ОТВЕТСТВЕННОСТИ.**

Взгляды, представленные в настоящем техническом директивном документе, отражают точку зрения авторов и не обязательно выражают позицию членов Исполнительного совета МВФ, руководства МВФ или любого из его государств-членов.

Рекомендуемая ссылка:

Aslett, Joshua, Gustavo González, Stuart Hamilton, and Miguel Pecho. 2024. "Tax Administration: Essential Analytics for Compliance Risk Management." IMF Technical Notes and Manuals 2024/01. International Monetary Fund, Washington, DC.

Просьба направлять заказы по адресу:  
International Monetary Fund, Publication Services PO  
Box 92780, Washington, DC 20090, USA  
Телефон: (202) 623-7430 | Факс: (202) 623-7201  
publications@IMF.org  
elibrary.IMF.org  
bookstore.IMF.org

# Содержание

Список сокращений.....	v
I. Введение .....	1
II. Теория, основы и принципы УКР.....	3
III. Понятия статистики и науки о данных.....	6
IV. Поддержка деятельности в области УКР: данные, технологии и инструменты.....	10
V. Поддержка деятельности в области УКР: основные аналитические процессы.....	15
<b>Приложения</b>	
Приложение 1. Доступ к сопроводительному инструментарию .....	25
Приложение 2. Необходимые профильные знания .....	26
Приложение 3. Решение ключевых задач: качество данных и кадровый потенциал .....	31
Приложение 4. Отдельная тема. Комплаенс-планирование .....	36
Приложение 5. Отдельная тема. Профилирование налогоплательщиков.....	39
Приложение 6. Отдельная тема. Отбор кандидатов для проведения проверок .....	43
<b>Литература.....</b>	<b>49</b>



## Список сокращений

БЯМ	Большая языковая модель
ИИ	Искусственный интеллект
ИТ	Информационные технологии
МДР	Модель дифференциации рисков
МКД	Межквартильный диапазон
НДС	Налог на добавленную стоимость
ОЭСР	Организация экономического сотрудничества и развития
РАД	Разведочный анализ данных
СФЛ	Состоятельное физическое лицо
УКР	Управление комплаенс-рисками
B2B	Взаимодействие между предприятиями
B2C	Взаимодействие между предприятием и потребителем
BISEP	Компания, отрасль, социологические, экономические, психологические факторы
CBC	По отдельным странам
ETL	Извлечение, преобразование, загрузка
G2B	Взаимодействие между государством и предприятием
G2G	Взаимодействие между государственными органами
NLP	Обработка естественного языка
OLAP	Оперативная аналитическая обработка данных
ROC-кривая	Кривая соотношения истинных и ложных характеристик
RPA	Роботизация
SAF-T	Стандартный файл налогового аудита
SQL	Язык структурированных запросов
VITARA	Виртуальное обучение для совершенствования работы в области государственных поступлений





# I. Введение

Настоящая справка предназначена для использования в качестве стартового набора по развитию практических аналитических возможностей, повышающих эффективность управления комплаенс-рисками (УКР) в рамках налогового администрирования. В начале справки приводится описание современной практики УКР с выделением принципов, которые влияют на использование данных. Далее изложен краткий обзор статистики и науки о данных как основных дисциплин, на которые опирается аналитика. В контексте УКР в справке представлены иллюстративные обзоры: 1) типичных данных, инструментов и технологий, а также 2) функций и результатов применения аналитики. В приложениях к справке рассматриваются важные области знаний, ключевые задачи и три отдельные темы (комплаенс-планирование, профилирование налогоплательщиков и отбор кандидатов для проведения проверок), каждая из которых имеет важное значение для УКР, но часто вызывает трудности в работе аналитических групп.

На Портале государственных доходов Департамента по бюджетным вопросам МВФ можно скачать сопроводительный инструментарий, содержащий дополнительные формы и инструменты<sup>1</sup>. Налоговые ведомства по всему миру используют множество разных аналитических технологий<sup>2</sup>. В соответствии с принятой политикой, МВФ не поддерживает какие-либо конкретные продукты или поставщиков. Если формы, представленные в сопроводительном инструментарии, будут приняты для использования в работе, их следует адаптировать для использования тех аналитических технологий, которые в наибольшей степени отвечают потребностям конкретного ведомства. Хотя в этих формах представлена современная передовая практика, они предоставляются исключительно в образовательных целях и без каких-либо гарантий. Инструкции по получению доступа к инструментарию и его использованию приводятся в приложении 1.

Рекомендации в настоящей справке основаны на широком определении аналитики как «практики использования математики для анализа данных». Независимо от применяемых методов, которые могут варьироваться от использования простой арифметики до сложных вычислительных алгоритмов, все формы аналитики опираются на математику. Поскольку математика универсально применима, аналитика потенциально может быть полезна в любой области, где существуют данные. Для ее применения необходимо сочетание технических навыков (например, в области статистики) и тематических знаний (например, по конкретной проблеме ведения бизнеса), позволяющих обрабатывать данные, обнаруживать закономерности, интерпретировать их и затем передавать их смысл. В настоящей справке эти и связанные с ними задачи рассматриваются как элементы, из которых состоит аналитика.

В налоговом администрировании данные используются повсеместно, а роль аналитики стала стратегически важной. Процесс сбора государственных доходов всегда опирался на информацию. Хотя концепции и методы со временем развивались, налоговые ведомства уже долгое время используют реестры, методы оценки и журналы для учета. В последние десятилетия эти идеи совершенствовались специалистами в сфере информационных технологий (ИТ), что способствовало развитию эффективных функций центральных ведомств, специализации персонала и способности использовать данные для принятия решений на всех уровнях. В связи с этим многие налоговые ведомства считают, что эффективность их аналитической работы оказывает воздействие на общие результаты их деятельности, особенно в том, что касается эффективного распределения ресурсов.

Помимо стратегической значимости аналитики, большинство налоговых администраций не могут эффективно выполнять свои основные задачи без ее использования. Примеры в подтверждение этому охватывают весь спектр административных функций. К числу важных функций, которые сегодня поддерживает

<sup>1</sup> Доступ по ссылке: <https://www.imf.org/revenueportal>.

<sup>2</sup> См. OECD, 2016.

аналитика, относятся выявление кандидатов на регистрацию, мониторинг подачи деклараций и уплаты налогов, исчисление налога, оценка объема просроченной задолженности, анализ сегментов налогоплательщиков, выявление случаев несоблюдения налоговых норм и прогнозирование будущего поведения налогоплательщиков.

**Поскольку аналитические возможности могут применяться в этих и многих других областях, УКР служит в качестве четкой призмы, через которую можно систематизировать их использование.** В следующих разделах аналитика рассматривается в контексте достаточно развитой практики УКР. Знакомство с ее возможностями начинается непосредственно с краткого обзора УКР.

## II. Теория, основы и принципы УКР

**УКР, как глубокая и постоянно развивающаяся область, в своей основе стремится влиять на поведение налогоплательщиков.** С концептуальной точки зрения ее логика основывается на теории и принципах социальных наук. При классической форме УКР специалисты стремятся оказывать воздействие посредством сочетания предоставления услуг и обеспечения соблюдения норм в соответствии с метафорой «кнута и пряника». При других формах УКР они стремятся укрепить аспекты доверия и социальных норм, существующих между отдельными налогоплательщиками и в их коллективных отношениях с государственными учреждениями. Независимо от выбранного подхода, эффективное использование УКР, как правило, в значительной степени опирается на анализ рисков и данные для содействия разработке адресных комплаенс-кампаний, стратегий и планов, подкрепленных продуманными тактическими интервенциями.

**Для ознакомления с полным набором базовых методов УКР рекомендуем использовать другие ресурсы, доступные в Интернете.** Специалисты, которые только начинают осваивать УКР, при желании могут изучить следующие ресурсы: 1) онлайн-курс по УКР, проводимый в рамках программы «Виртуальное обучение для совершенствования работы в области государственных поступлений» (VITARA) и 2) технические справки МВФ, описывающие УКР, внедрение основ УКР и разработку планов по улучшению соблюдения норм<sup>3</sup>.

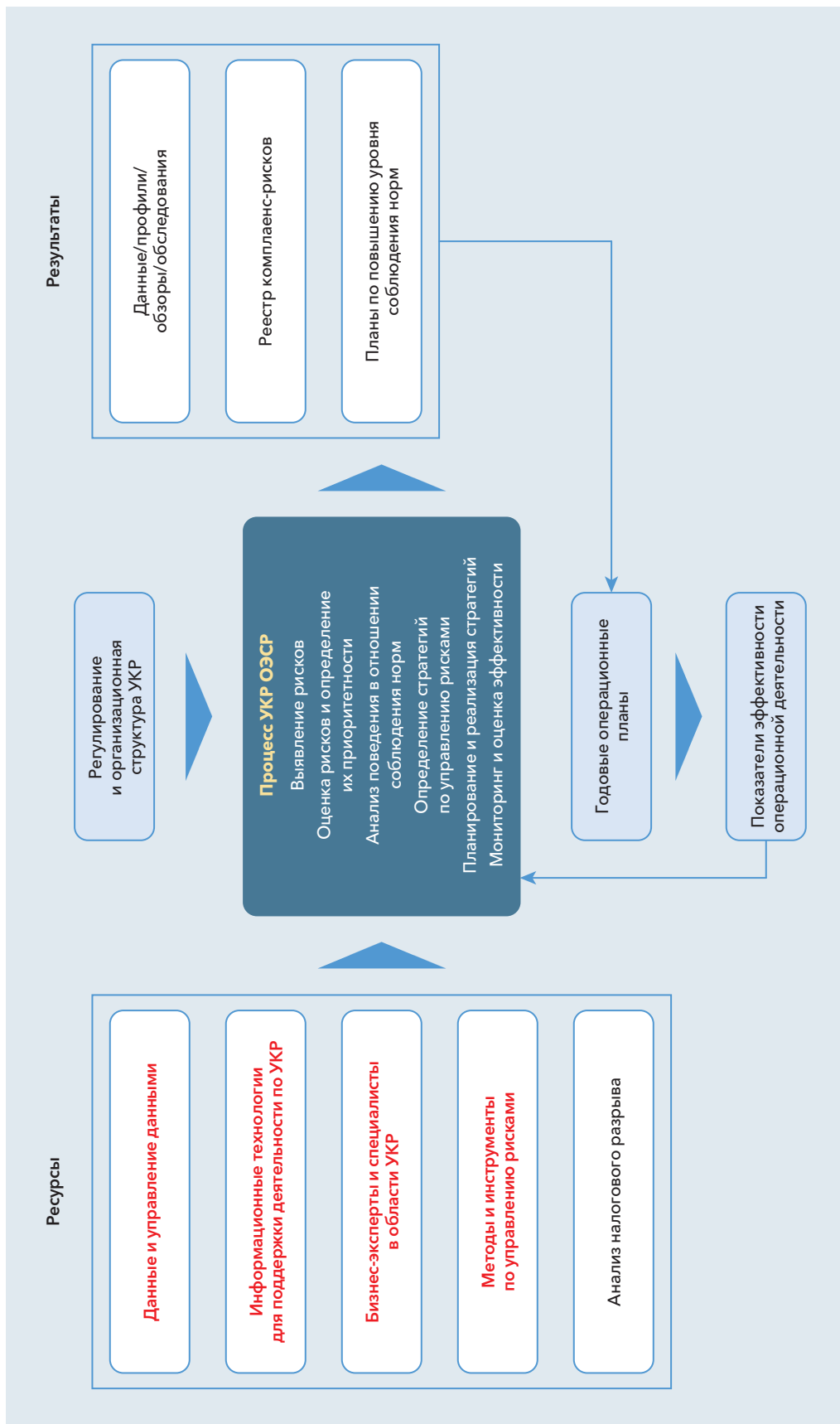
**В основе всех этих ресурсов лежит идея о том, что в каждой конкретной юрисдикции существует состояние равновесия квази-соблюдения налоговых норм, и УКР может влиять на его баланс.** Это равновесие, основанное на совокупности мнений, действий и убеждений отдельных лиц и ведомств, действующих в рамках одной налоговой системы, представляет собой статус-кво поведения налогоплательщиков. Результатом его баланса является разрыв соблюдения налоговых норм (то есть разность между потенциальными налоговыми поступлениями, которые могут быть получены в условиях действующего законодательства, и фактическими поступлениями). Величина таких разрывов, состоящих из разных компонентов, различается в зависимости от отрасли, сегмента и рассматриваемой проблемы. Многие ведомства считают сокращение разрывов важной стратегической задачей, в выполнении которой может помочь УКР. Для этого необходимы ресурсы, аналитические навыки более высокого порядка и понимание процессов УКР.

**Как показано на рис. 1, система УКР хорошо подходит для использования аналитики, поскольку объединяет данные и экспертные знания, позволяющие определять риски, интерпретировать поведение и принимать решения о проведении интервенций.** С концептуальной точки зрения внедрение процессов УКР, как правило, происходит с использованием процессуальной основы, опубликованной Организацией экономического сотрудничества и развития (ОЭСР) в 2004 году. Результаты применения этой основы позволяют получить целый ряд аналитических данных. С помощью методов анализа сотрудники групп по УКР определяют, оценивают, прогнозируют и предотвращают комплаенс-риски. Помимо описания рисков, они активно стремятся понять причины проявления конкретных форм поведения. Благодаря более глубокому анализу, помимо разовых мероприятий по обеспечению соблюдения норм, таких как аудиторские проверки, меры, принимаемые в ответ на риски, могут включать комплексные стратегии, направленные на решение проблем, носящих более структурный характер. На рис. 1 представлена разработанная ОЭСР основа с указанием ключевых исходных ресурсов и результатов.

**Основное внимание в настоящей справке уделяется элементам, выделенным на рис. 1 красным, и аналитической работе, которую они позволяют проводить.** Вопрос интегрированного применения ресурсов в системе УКР, вызывающий особый интерес во многих администрациях, может быть непрозрачной областью. Во многих юрисдикциях информация о конкретных используемых средствах и методах является конфиденциальной и редко раскрывается. В других юрисдикциях могут публиковаться определенные детали в разных

<sup>3</sup> См. Chooi and Pemberton (2023), Betts (2022) и Brondolo et al. (2022).

Рисунок 1. Основные элементы системы УКР



Источники: Betts, 2022; персонал МВФ.  
 Примечание. ОЭСР – Организация экономического сотрудничества и развития; УКР – управление комплаенс-рисками.

целях (например, в целях демонстрации справедливости процессов или поддержки соответствующих инициатив по борьбе с коррупцией). Кроме того, между администрациями существуют значительные различия по уровню развития и сложности как самих исходных ресурсов, так и конкретных используемых аналитических методов. С учетом этих сложностей рекомендации в настоящей справке предназначены для иллюстрации, основанной на передовых методах, которые регулярно наблюдаются на практике.

**Прежде чем перейти к рассмотрению ресурсов и аналитических возможностей, важно отметить, что оба аспекта, как правило, определяются основным набором принципов УКР.** Через УКР налоговым ведомствам следует:

- **Начинать с основ.** В основе практики УКР лежат четыре компонента соблюдения налоговых норм: регистрация, подача деклараций, представление данных и уплата налогов. Хотя необходимость активного управления соблюдением норм по каждому компоненту существует всегда, при использовании аналитики компонент представления данных часто рекомендуется подробно рассматривать только после глубокого понимания компонентов регистрации, подачи деклараций и уплаты налогов.
- **Делать упор на добровольном соблюдении налоговых норм.** С учетом социальных аспектов налогового администрирования цель УКР – максимально повысить добровольное соблюдение налогоплательщиками законодательства при наименьших издержках. Интервенции в любой форме требуют ресурсов, причем в некоторых случаях более значительных, чем в других (аудит), поэтому способность воздействовать на макроуровне на поведение налогоплательщиков по каждому из четырех компонентов без необходимости высоких затрат имеет крайне важное значение.
- **Проводить различие между разными категориями налогоплательщиков и проблем.** Поскольку характер и степень соблюдения норм могут различаться в зависимости от сегмента, отрасли и существующих проблем, недифференцированный подход к анализу рисков редко бывает целесообразным. Эта логика также справедлива и для комплаенс-планирования, при котором требуется надлежащая адаптация проводимых в любой форме интервенций.
- **Интегрировать применение административных инструментов.** Комплаенс-кампании, стратегии и планы должны учитывать весь спектр потенциально доступных ресурсов для проведения комплексных интервенций, которые в совокупности будут обеспечивать максимальное воздействие (например, сочетание сопоставления данных, освещения в СМИ, предоставления услуг, проведения аудиторских проверок и передачи дел в суд).
- **Уделять приоритетное внимание изменению текущего поведения налогоплательщиков.** При подходе к УКР на основе реагирования усилия традиционно концентрируются на отборе кандидатов для проведения проверок на основе анализа решений, принимавшихся налогоплательщиками в прошлом (отраженных в налоговых декларациях и других данных). Несмотря на важность этого подхода, воздействуя на решения, принимаемые в реальном времени, можно достичь более значительных результатов.
- **Применять подход, ориентированный на получение новых данных.** Сразу после составления данных они становятся отражением прошлого. В связи с этим при осуществлении УКР, особенно в тех юрисдикциях, которые испытывают трудности с управлением данными, налоговым администрациям следует в значительной степени опираться на информацию, представляемую экспертами, аналитиками, партнерскими ведомствами, информаторами и другими ресурсами на местах.

**С учетом этих принципов изучение аналитических возможностей для УКР также требует понимания потенциала статистики и науки о данных.** Эти две области являются неотъемлемой частью всех аспектов современной аналитики, особенно тех из них, которые по своему характеру являются прогнозными и используются все чаще. Несмотря на то, что доступные сегодня инструменты значительно сокращают потребность в понимании математических методов, на которых они основаны, рабочее понимание базовой полезности этих двух областей все же необходимо.

### III. Понятия статистики и науки о данных

**Статистика – это область математической науки, направленная на описание и интерпретацию данных с помощью методов, которые позволяют измерять показатели вероятности и неопределенности.** Использование статистики является краеугольным камнем самого научного подхода. В классических подходах к статистике, лежащих в основе современной аналитики, как правило, используются сравнительно небольшие, но тщательно проработанные наборы данных. В связи с меньшими объемами данных, статистика традиционно используется для моделирования экспериментов, проверки гипотез и установления взаимосвязей между переменными. Эти методы часто применяются для определения того, в какой степени выводы могут быть получены и обоснованы. Для того чтобы выполнять эту работу, профессиональные статистики часто имеют образование непосредственно в области статистики, прикладной математики или тесно связанной с ними дисциплины.

**Для сравнения, наука о данных – это относительно новая междисциплинарная область, которая опирается на статистику, информатику и бизнес.** Наука о данных, возникшая как следствие развития современной вычислительной техники и ИТ, все чаще признается в качестве фундаментальной области знаний или научной дисциплины. Специалисты по обработке и анализу данных работают с относительно большими объемами данных, используя те же принципы, которыми пользуются статистики. Однако, опираясь на основы, заложенные статистикой, они, как правило, уделяют основное внимание более операционным аспектам деятельности организаций, которые они поддерживают, занимаясь разработкой продуктов интеллектуальной обработки данных, интеллектуальным анализом данных, написанием программного обеспечения и интеграцией результатов своей работы в бизнес-системы – специализация, которая хорошо сочетается с исходными ресурсами в системе УКР, представленными на рис. 1.

**Хотя наука о данных строится на базе статистики, между этими двумя областями трудно провести четкую грань, которая представляет собой спорный вопрос.** На рис. 2 показаны области, в которых наблюдаются интеграция и пересечение. Смешение дисциплин не только не вызывает размывания ни одной из областей, но также приводит к разработке новых, передовых способов применения уже известных идей (в частности, машинного обучения, которое сегодня является одним из основных направлений работы в области искусственного

Рисунок 2. Области знаний: статистика и наука о данных



Источник: персонал МВФ.

интеллекта [ИИ]). В центре этих изменений – наука о данных, которая стала движущей силой стремительных инноваций. Базовые дисциплины, на которые она опирается, также продолжают развиваться, но их развитие, как правило, сопровождается повышенными требованиями к точности и поэтому происходит медленнее.

**Освоение навыков в области статистики и науки о данных требует понимания терминологии, основных концепций и их применения.** Во вставках 1 и 2 приводится их обзор. Эти детали используются во всех областях, указанных на рис. 2, с некоторыми вариациями.

### ВСТАВКА 1. Основные понятия

**Данные.** Факты, значения, показатели или наблюдения, получаемые и используемые в качестве основы для умозаключений, анализа, изучения или расчетов.

**Генеральная совокупность.** Совокупность сходных организаций, лиц, объектов или событий, отобранных для исследования и имеющих по меньшей мере одну общую характеристику.

**Выборка.** Часть более крупной генеральной совокупности, отобранная с использованием предварительно определенного метода с целью обеспечить возможность проведения исследования и анализа.

**Функция.** Математическое выражение или набор правил, определяющий взаимосвязи между набором переменных, часто представленный в виде формулы.

**Распределение.** Набор данных по конкретной переменной (например, по размеру задекларированного дохода), который может соответствовать математической функции.

**Доверительный интервал.** Показатель неопределенности, часто интерпретируемый как вероятность того, что в результате эксперимента будут получены схожие результаты в определенном диапазоне значений.

**Алгоритм.** Логика, методы или наборы конкретных шагов для решения проблемы, которые могут быть кодифицированы для использования на компьютерном языке программирования.

**Искусственный интеллект.** Моделирование человеческого интеллекта, осуществляемое с помощью компьютерных систем и программного обеспечения, в значительной степени опирающееся на статистику и науку о данных.

**Машинное обучение.** Одна из основных областей искусственного интеллекта, предполагающая использование данных и алгоритмов, позволяющих обучать компьютерные системы без непосредственного программирования.

**Контролируемое обучение.** Категория машинного обучения, при которой специалист помечает данные и обучает алгоритмы делать выводы или прогнозы.

**Неконтролируемое обучение.** Категория машинного обучения, которая не требует обучения человеком для выявления закономерностей в данных или получения выводов или прогнозов.

**Интеллектуальный анализ данных.** Использование статистики и науки о данных для анализа больших наборов данных и выявления аномалий, корреляций, закономерностей и тенденций.

**Структурированные данные.** Данные с четким определением взаимосвязей и типов данных (например, нормализованные таблицы базы данных), по которым можно легко осуществлять поиск.

**Неструктурированные данные.** Данные, не имеющие четкой структуры, часто хранящиеся в форматах, требующих использования специализированных инструментов для анализа (например, счета-фактуры, отсканированные в виде изображений).

## ВСТАВКА 2. Основные концепции и потенциальная полезность

### Современная аналитика строится на базе статистики:

- **Описательная статистика** описывает или обобщает характеристики и распределение набора данных, используя такие понятия, как среднее значение, медиана, мода, диапазон, перекокс и дисперсия.
- **Индуктивная статистика** применяет высшую математику для создания четких, структурированных моделей, которые позволяют изучать данные выборки, проверять гипотезы и делать выводы о более крупной генеральной совокупности.
- **Прогнозная аналитика** опирается на статистику и алгоритмы машинного обучения для анализа часто значительных объемов данных и прогнозирования вероятных будущих результатов.
- **Предписывающая аналитика** дополняет выводы прогнозной аналитики за счет анализа результатов предыдущих действий и рекомендаций в отношении дальнейших оптимальных действий.

### Статистика и наука о данных позволяют получать результаты, имеющие практическую ценность:

- **Классификация наблюдений** с использованием бинарных или многоклассовых методов прогнозирования (например, бинарный метод: «соблюдает нормы» или «не соблюдает нормы»; многоклассовый метод: «высокий», «средний» или «низкий» риск).
- **Оценка значений** с использованием линейных и нелинейных оценок для прогнозирования выходных переменных (например, сумм корректировок по результатам аудиторской проверки).
- **Оценка вероятностей** с использованием различных методов (например, алгоритма случайного леса) для перехода от использования балльных оценок риска к более точным вероятностям риска.
- **Группировка наблюдений** с использованием неконтролируемых методов (например, самоорганизующихся карт) для выявления и визуального представления общих свойств, которые не были обозначены или которые сразу не очевидны.
- **Анализ сетей** с использованием специализированных инструментов для установления связей между объектами, операциями и событиями в больших объемах структурированных и неструктурированных данных.

### Оценка эффективности имеет важнейшее значение и начинается с двух понятий:

- **Матрица несоответствий**, применяемая для обобщения показателей верных и ошибочных прогнозов алгоритма классификации, представляет собой фундаментальное понятие науки о данных.
- **Кривая соотношения истинных и ложных характеристик** применяется для визуального представления и подтверждения прогнозирующих возможностей алгоритма до его использования на практике.

### Машинное обучение становится все более актуальным:

Способность алгоритмов машинного обучения обнаруживать закономерности в больших наборах данных коренным образом меняют практику оценки рисков. Традиционно аналитики в значительной степени полагались на экспертные правила и детальные статистические модели, которые позволяют анализировать риски по методу «сверху вниз» с использованием отдельных показателей в рамках отраслей и сегментов налогоплательщиков. Эффективность этих методов во многом зависела от знаний и навыков соответствующих специалистов. Для сравнения, большинство базовых алгоритмов машинного обучения (например, деревья принятия решений) позволяют с высокой степенью точности устанавливать взаимосвязи в данных, связанных с рисками, и делать прогнозы по методу «снизу вверх».



**В сфере налогового администрирования ресурсы УКР, к которым применяются эти концепции, включают в себя широкий набор данных, технологий и инструментов.** Несмотря на значительные различия, в их использовании существуют устойчивые закономерности. При рассмотрении вопроса о внедрении новых возможностей важно иметь общее представление об этих деталях.

## IV. Поддержка деятельности в области УКР: данные, технологии и инструменты

УКР как процесс опирается на поддерживающие его учреждения, многие из которых инвестируют значительные средства в цифровые технологии и реформы. Ресурсы в системе УКР, как правило, отражают уровень развитости создающих их учреждений, эффективность их деятельности и кадровый потенциал в целом. Инвестиции в ИТ часто оказывают преобразующее воздействие на все эти области. Они позволяют проводить структурные реформы, направленные на централизацию информации, отказ от использования бумажных носителей, укрепление функций головного офиса, реорганизацию процессов, а также значительное повышение эффективности сбора данных, управления ими и их использования. Поскольку на сегодняшний день почти все налоговые администрации автоматизировали основные аспекты своей деятельности, в настоящее время многие из них пользуются более широким спектром цифровых возможностей.

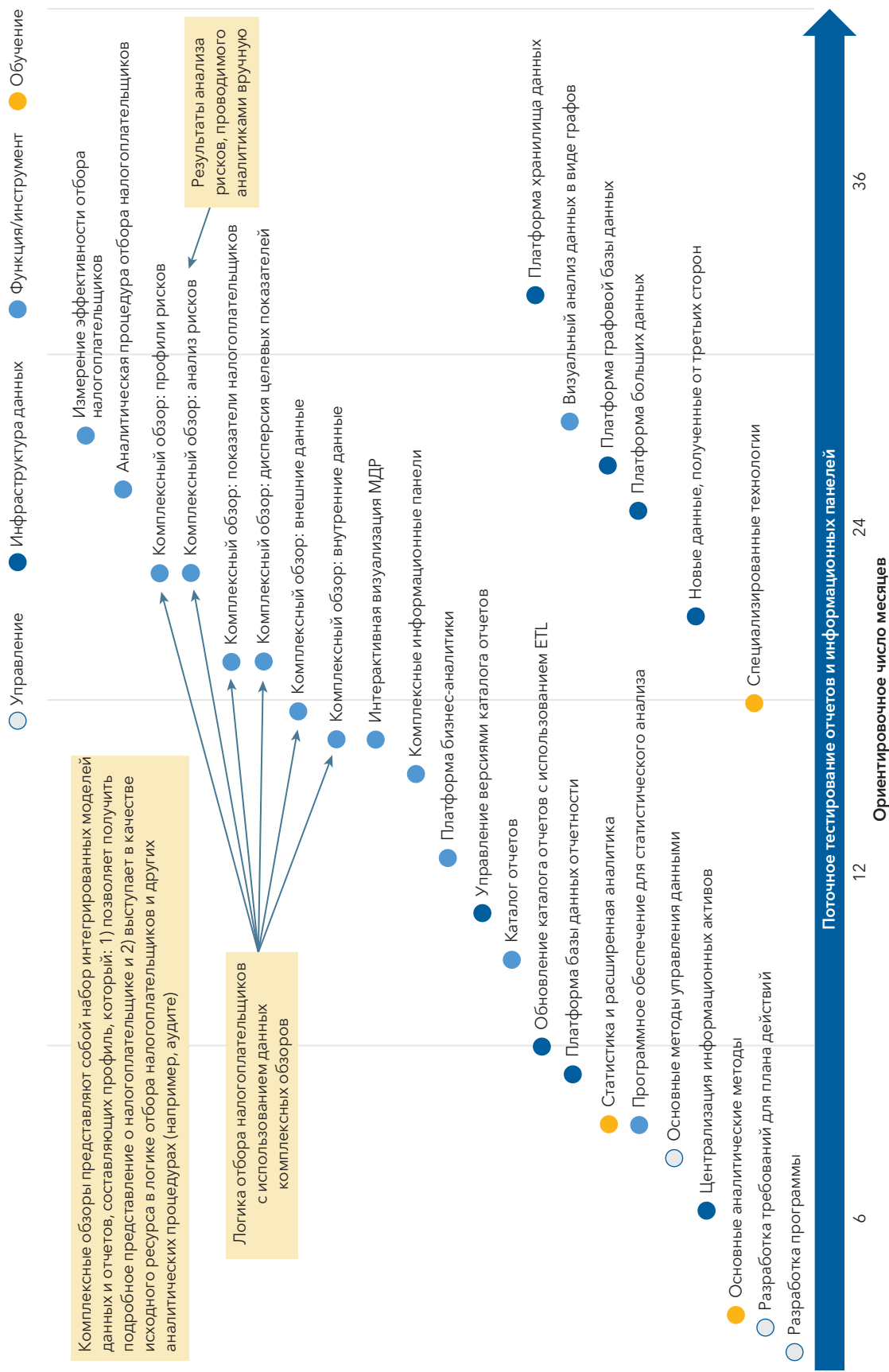
Для получения дополнительной информации о цифровизации, данных и информационных технологиях в целом можно обратиться к другим ресурсам, доступным в Интернете. К их числу относятся: 1) готовящийся к публикации курс VITARA по ИТ и управлению данными и 2) технические справки МВФ по стратегическому планированию развития ИТ в сфере налогового администрирования, основным ИТ-системам для налогового администрирования и внедрению коммерческих готовых продуктов для базовых систем<sup>4</sup>.

Среди изменений, возможных в налоговом администрировании благодаря цифровизации, процесс УКР уникален тем, что его технологические ресурсы, как правило, не поддаются аккуратному объединению в единую систему. По ряду практических причин это может усложнять их развитие. Так, в случае одного из ключевых примеров, закупок для УКР, могут возникать более значительные трудности по сравнению с другими областями в связи с возможным наличием нескольких поставщиков. Использование нескольких разных продуктов и ИТ-платформ является обычной практикой, а их интеграция и развитие, как правило, требуют активного участия специалистов из различных областей, в том числе управления проектами, разработки программного обеспечения, разработки баз данных, ИТ-администрирования и операционной деятельности.

Для содействия в изучении этих вопросов на рис. 3 представлен иллюстративный обзор типичных, ориентированных на использование технологий ресурсов в системе УКР в развитой налоговой администрации. Более подробное описание всех ресурсов, визуально представленных в виде «дорожной карты», приводится в таблице 1. При рассмотрении рис. 3 обратите внимание на категории ресурсов и их последовательное представление. Разработка совокупности представленных элементов часто является более подходящей отправной точкой для определения масштабов проекта по развитию потенциала в области УКР, чем концентрация усилий на отдельной системе рисков. Элементы «дорожной карты», которые на сегодняшний день необходимо иметь в большинстве налоговых администраций, отмечены в таблице 1 галочкой как необходимые. Конкретно в области УКР большинство используемых технологических ресурсов создаются на основе слоев технологий и инфраструктуры данных, внедряемых с течением времени.

<sup>4</sup> См. Aslett et al. (готовится к публикации) и Cotton and Dark (2017a, 2017b, 2017c).

**Рисунок 3. Иллюстрация плана действий по поддержке деятельности в области УКР: данные, технологии и инструменты**



Источник: персонал МВФ.  
 Примечание. ETL – процессы извлечения, преобразования и загрузки; МДР – модель дифференциации рисков; УКР – управление комплаенс-рисками.

Таблица 1. Иллюстрация плана действий по поддержке деятельности в области УКР: данные, технологии и инструменты

Необходимо? Основные этапы		Примечания	
Управление	1 ✓	Разработка программы	«Дорожную карту» следует реализовывать как совокупность проектов.
	2 ✓	Разработка требований для плана действий	Следует создавать документы с формальными требованиями и управлять ими.
	3 ✓	Основные методы управления данными	Необходимы стратегии в отношении жизненного цикла данных, архитектура данных и словари данных.
	4 ✓	Централизация информационных активов	Централизация всех рабочих наборов данных путем консолидации платформ или репликации.
	5 ✓	Платформа базы данных отчетности	Временная платформа для отчетов по запросу и специального анализа.
	6 ✓	Обновление каталога отчетов с использованием ETL	Автоматизация пакетов операций «извлечение, преобразование, загрузка» (ETL).
Инфраструктура данных	7	Управление версиями каталога отчетов	Обеспечение возможности поиска данных в ключевых отчетах на определенный момент времени.
	8	Новые данные, полученные от третьих сторон	Получение и использование внешних данных при условии соблюдения надлежащих мер контроля.
	9	Платформа больших данных	Автономная платформа, часто используемая в качестве «озера данных» для массового хранения данных.
	10	Платформа графовой базы данных	Платформа базы данных, предназначенная для поддержки анализа сетей.
	11	Платформа хранилища данных	Внедрение традиционного хранилища данных – крупный проект.
Функция/инструмент	12 ✓	Программное обеспечение для статистического анализа	Используется как для статистического анализа, так и для прогнозной аналитики.
	13 ✓	Каталог отчетов	Обеспечивает доступ по требованию к каталогу стандартизированных отчетов.
	14	Платформа бизнес-аналитики	Облегчает публикацию информационных панелей, отчетов и аналитических материалов.
	15 ✓	Комплексные информационные панели	Панели с информацией о проводимой работе и соблюдении норм.
	16	Интерактивная визуализация МДР	Инструмент управления рисками и планирования для анализа сегментов налогоплательщиков.
	17 ✓	Комплексный обзор: внутренние данные	Позволяет получить обзор внутренних данных налоговой администрации о налогоплательщике на определенный момент времени.

Таблица 1. Иллюстрация плана действий по поддержке деятельности в области УКР: данные, технологии и инструменты (окончание)

		Необходимо?	Основные этапы	Примечания
Функция/инструмент	18	✓	Комплексный обзор: внешние данные	Позволяет получить обзор данных о налогоплательщике из внешних источников на определенный момент времени.
	19	✓	Комплексный обзор: дисперсия целевых показателей	Показывает показатели налогоплательщика относительно отдельных контрольных показателей.
	20	✓	Комплексный обзор: показатели налогоплательщиков	Показывает налоговые, финансовые и другие важные коэффициенты, рассчитанные для налогоплательщика.
	21	✓	Комплексный обзор: анализ рисков	Показывает результаты анализа рисков, проводимого специалистами вручную.
	22	✓	Комплексный обзор: профили рисков	Показывает результаты анализа рисков и предложения по дальнейшим действиям.
	23	✓	Аналитическая процедура отбора налогоплательщиков	Уровни анализа с использованием алгоритмов и интеллектуального анализа данных.
	24	✓	Измерение эффективности отбора налогоплательщиков	Набор показателей, позволяющих отслеживать эффективность отбора налогоплательщиков для проверок.
	25		Визуальный анализ графических данных	Интерактивные инструменты, использующие графовую базу данных для анализа сетей.
Обучение	26	✓	Основные аналитические методы	Темы: электронные таблицы, поиск по базе данных и разработка отчетов.
	27		Статистика и расширенная аналитика	Темы: описательная статистика, индуктивная статистика и прогнозная аналитика.
	28		Специализированные технологии	Темы: хранилище данных, большие данные и платформы графовых баз данных.

Источник: персонал МВФ.

Помимо элементов, указанных на рис. 3 и в таблице 1, важное значение могут иметь следующие дополнительные инструменты и возможности, которые активно используются налоговыми администрациями в разных регионах мира:

- **Реестры комплаенс-рисков.** В развитой системе группы специалистов по УКР должны иметь возможность в любой момент времени представлять данные об основных комплаенс-рисках в своих юрисдикциях и о мерах, принимаемых для их снижения. Этому способствует наличие централизованного инструмента для учета и управления этими рисками.
- **Профили рисков по отраслям.** Концептуально схожий с комплексным обзором налогоплательщика, описанным в таблице 1, отраслевой профиль имеет как количественные, так и качественные аспекты. На практике специалист по анализу рисков, составляющий профиль рисков для отрасли, отражает соответствующие данные в базе данных о рисках. В случае интеграции систем в профили рисков налогоплательщиков включаются соответствующие риски из отрасли, к которой они относятся.
- **Профили рисков по операциям.** Помимо профилирования налогоплательщиков и отраслей, все больший интерес вызывает профилирование отдельных операций налогоплательщиков с использованием аналогичных концепций. В случаях электронного выставления счетов-фактур этот вопрос особенно актуален с точки зрения стандартов отчетности (например, стандартного файла налогового аудита).
- **Обработка данных о рисках в режиме реального времени.** Аналитика рисков применяется в режиме реального времени для оптимизации и распределения задач (например, для оценки рисков по счетам-фактурам в рамках логической системы путем применения надлежащего набора средств контроля и отбора счетов-фактур для ручной проверки или принятия мер) с использованием специализированных технологий, особенно в случае больших объемов операций.
- **Интеграция платформ.** В последние годы некоторые поставщики проводят интеграцию своих инструментов аналитики с другими продуктами (например, интеграцию программ для статистической обработки данных с платформами для управления рабочими процессами и данными о проверках). Создаваемые в результате этого возможности способны ускорить важные аспекты преобразований в области налогового администрирования, особенно при работе с налогоплательщиками.
- **Встроенная аналитика.** Многие налоговые администрации интегрируют аналитику рисков в ключевые рабочие процессы непосредственно на основных платформах обработки данных. В отношении разных процессов, таких как логика отбора кандидатов для проведения комплаенс-проверок, производство возвратов и непосредственно регистрация, аналитики часто вносят обновления путем изменения конфигурации и программирования в соответствующих программах и базах данных.
- **Роботизация (RPA).** Роботизация, являясь новой темой, представляющей интерес, создает широкие возможности в сфере налогового администрирования благодаря способности сочетать аналитику и особенно машинное обучение с использованием ботов, имитирующих человеческую деятельность. Практическое применение этой технологии в налоговом администрировании потенциально может иметь большое значение.

В основе инфраструктуры данных, технологий и инструментов, которые могут содействовать деятельности в области УКР, лежат сами данные, и диапазон их потенциального применения очень значителен. Для получения представления о потенциале аналитики в налоговом администрировании полезно изучить ее типичные результаты и функции.

## V. Поддержка деятельности в области УКР: основные аналитические процессы

При рассмотрении возможных аналитических процессов через призму УКР следует изучить вопросы о том, как их организуют, применяют и понимают специалисты. Во вставке 3 представлена модель их организации. Она наиболее целесообразна для групп специалистов по аналитике в развитых налоговых

### ВСТАВКА 3. Организация аналитической работы для поддержки УКР

#### Комплаенс-стратегия и планирование

1. Сквозные вопросы
2. Сегменты и проблемы
  - a. Крупные, средние, мелкие налогоплательщики
  - b. Секторы и отрасли
  - c. Категории рисков
  - d. Физические лица

#### Активные меры предотвращения

1. Прогнозирование соблюдения норм
2. Адресная работа с клиентами
  - a. Следующие оптимальные шаги
3. Встроенная аналитика
4. Аналитика в режиме реального времени
5. Сотрудничество с третьими сторонами

#### Мониторинг и обнаружение

1. Мониторинг основных вопросов соблюдения норм
2. Автоматическое обнаружение занижения данных
3. Разведочный анализ данных
4. Аналитика профилей
5. Сотрудничество с третьими сторонами

#### Обеспечение соблюдения налоговых норм

1. Отбор налогоплательщиков для проверок соблюдения норм
  - a. Следующие оптимальные шаги
2. Эффективность отбора налогоплательщиков для проверок
3. Оценка и уведомления
4. Сотрудничество с третьими сторонами

#### Макростратегический анализ и таргетирование

Благодаря использованию различных методов и специализированных моделей результаты позволяют поддерживать комплаенс-кампании, стратегии, операционные планы и рутинную деятельность специалистов по анализу рисков и обеспечивают необходимую для этого информацию.

#### Аналитика для превентивных интервенций

Результаты и модели, при работе с которыми используются все виды аналитики, от описательной до предписывающей, предоставляются рабочим группам для принятия мер и интегрируются в ИТ-сервисы.

#### Аналитика для исследований и операционной деятельности

Результаты, для работы с которыми требуются глубокое понимание описательной статистики и навыки работы с базовыми наборами данных, обеспечивают необходимую специалистам по анализу рисков информацию и обеспечивают работу инструментов (комплексные обзоры).

#### Аналитика для принятия ответных мер

Результаты, при работе с которыми также используются все виды аналитики, повышают эффективность работы с налогоплательщиками и поддерживают деятельность по обеспечению соблюдения норм и проведение мер, принимаемых в сотрудничестве со сторонними ведомствами.

Источник: персонал МВФ.

Примечание. ИТ – информационные технологии; УКР – управление комплаенс-рисками.

администрациях, но может быть полезна в качестве целевого ориентира и другим. Ее элементы используются при определении стратегии, планировании и в повседневной деятельности.

**В качестве целевого ориентира в таблицах 2-5 представлены примеры функций и результатов применения аналитики по всем вопросам, перечисленным во вставке 3.** Элементы, которые на сегодняшний день следует иметь большинству налоговых администраций, помечены как «необходимые». Для содействия в понимании процесса разработки каждого элемента для них указан основной вид используемой аналитики: «описательная», «прогнозная» или «предписывающая». В случае применения более одного вида аналитики используется обозначение «гибридная». В доступном для скачивания инструментарии, разработанном в дополнение к настоящей справке, приводятся функциональные примеры некоторых из описанных результатов и функций. Инструкции по получению доступа к инструментарию представлены в приложении 1.

**Поскольку настоящая справка представляет собой стартовый набор, ее сопровождают дополнительные ресурсы и справочные материалы, представленные в следующих шести приложениях.** Ниже приводится описание приложений, предназначенных для практического изучения функций и результатов применения аналитики, представленных в таблицах 2-5:

- **Приложение 1. Доступ к сопроводительному инструментарию.** Инструкции по загрузке и использованию сопровождающего эту справку инструментария, в который входит ряд разработанных инструментов и форм, настроенных на использование синтетических (искусственных) наборов данных, предназначенных главным образом для образовательных целей.
- **Приложение 2. Необходимые профильные знания.** Информация по следующим темам: 1) общие подходы к использованию данных в целях выявления налогоплательщиков, представляющих интерес; 2) важные статистические распределения для анализа налоговых данных; 3) использование разведочного анализа данных для выявления аномалий и резко отклоняющихся значений в данных; 4) новые разработки в области использования ИИ.
- **Приложение 3. Решение ключевых задач: качество данных и кадровый потенциал.** Общие подходы к решению проблем, связанных с качеством данных и кадровым потенциалом, эффективность которых отмечается налоговыми администрациями в разных регионах мира.
- **Приложение 4. Отдельная тема. Комплаенс-планирование.** Краткий концептуальный обзор комплаенс-планирования и обзор модели дифференциации рисков (МДР), полезной для проведения макростратегического анализа рисков с целью определения направления деятельности (стратегий и планов) по обеспечению соблюдения налоговых норм. Включает иллюстрацию результатов использования формы МДР на основе Excel, представленной в сопровождающей справке инструментарию.
- **Приложение 5. Отдельная тема. Профилирование налогоплательщиков.** С целью подчеркнуть важность составления профилей налогоплательщиков вручную (особенно для крупных налогоплательщиков или в других сложных случаях) представлен подход по разработке цифровых инструментов для содействия анализу рисков. Включает иллюстрацию результатов использования формы профилирования налогоплательщиков на основе Excel, представленной в сопровождающей справке инструментарию.
- **Приложение 6. Отдельная тема. Отбор кандидатов для проведения проверок.** Представляет современные концепции, влияющие на использование аналитики при отборе кандидатов для проверок, с подчеркиванием особой важности использования сочетания правил оценки рисков, разработанных профильными экспертами, и прогнозной аналитики с использованием алгоритмов. Включает иллюстрацию результатов использования рабочего процесса на основе прогнозной аналитики, представленного в сопровождающей справке инструментарию.



Таблица 2. Пример использования аналитики: комплаенс-стратегия и планирование

Необходимо? Вид аналитики		Результат или функция	Примечания
<b>1</b>	<b>Сквозные вопросы</b>		
1.1	✓ Описательная	Формы для определения стратегий и планирования	Предоставление форм, предварительно заполненных соответствующими статистическими данными
1.2	Описательная	Анализ охвата деятельности по обеспечению соблюдения норм	Выявление пробелов в охвате стратегий и планов
1.3	Описательная	Анализ эффективности: стратегии и планы	Оценка воздействия стратегий и планов на макроуровне
1.4	Описательная	Анализ эффективности: управляемые рабочие процессы	Оценка деятельности, способствующей реализации стратегий и планов
1.5	Гибридная	Кластерный анализ для сегментации	Выявление неизвестных способов сегментации налогоплательщиков
1.6	✓ Описательная	Кумулятивный метод расчета: оборот и налог	Определение наиболее важных налогоплательщиков в сегменте
1.7	✓ Гибридная	Определение и выбор задач	Регулярный отбор налогоплательщиков, объектов и событий, в отношении которых требуется принятие мер
<b>2</b>	<b>Сегменты и проблемы</b>		
<b>2.a.i</b>	<b>Крупные налогоплательщики</b>		
2.a.i.1	✓ Описательная	Пересмотр критериев отнесения налогоплательщиков к крупным	Предложение вариантов уточнения существующих критериев
2.a.i.2	✓ Описательная	Анализ МДР: все налогоплательщики	Оценка всех крупных налогоплательщиков в рамках единого анализа
2.a.i.3	Описательная	Анализ МДР: бухгалтеры	Оценка бухгалтеров по их агрегированной клиентской отчетности
<b>2.a.ii</b>	<b>Средние налогоплательщики</b>		
2.a.ii.1	✓ Описательная	Анализ МДР: наиболее важные налогоплательщики	Оценка только наиболее крупных налогоплательщиков средней категории (например, ~1 000)
2.a.ii.2	Описательная	Анализ МДР: Таможенные брокеры	Оценка брокеров по их агрегированной клиентской отчетности
<b>2.a.iii</b>	<b>Мелкие налогоплательщики</b>		
2.a.iii.1	✓ Описательная	Анализ пороговых значений: концентрация налогоплательщиков на пороговых уровнях	Анализ всех ключевых пороговых значений (например, НДС)
<b>2.b</b>	<b>Секторы и отрасли</b>		
2.b.1	✓ Описательная	Анализ МДР: наиболее важные налогоплательщики	Оценка только наиболее крупных налогоплательщиков по отраслям (например, ~1 000)
2.b.2	Индуктивная	Модель БАДО: горнодобывающая промышленность	Анализ минеральных ресурсов для учета при планировании
2.b.3	Индуктивная	Модель БАДО: нефть и газ	Анализ нефтегазовых ресурсов для учета при планировании

(продолжение)

Таблица 2. Пример использования аналитики: комплаенс-стратегия и планирование (окончание)

Необходимо? Вид аналитики		Результат или функция	Примечания
<b>2.c</b>	<b>Категории рисков</b>		
2.c.1	Гибридная	Специальные исследования по ключевым темам	Например, перераспределение прибыли, трансфертное ценообразование, мошенничество в сфере НДС
<b>2.d</b>	<b>Физические лица</b>		
2.d.1	✓ Описательная	Пересмотр критериев отнесения налогоплательщиков к числу СФЛ	Предложение вариантов уточнения существующих критериев
2.d.2	✓ Описательная	Анализ МДР: все налогоплательщики, относящиеся к СФЛ	Анализ всех налогоплательщиков, относящихся к СФЛ, в рамках единого анализа

Источник: персонал МВФ.

Примечание. БАДО – налогово-бюджетный анализ добывающих отраслей; МДР – модель дифференциации рисков; НДС – налог на добавленную стоимость; СФЛ – самостоятельное физическое лицо.

Таблица 3. Пример использования аналитики: активные меры предотвращения

Необходимо? Вид аналитики		Результат или функция	Примечания
<b>1</b>	<b>Прогнозирование соблюдения норм</b>		
1.1	Прогнозная	Будущая регистрация: прогнозирование нерегистрации	Кого будет необходимо зарегистрировать, с последствиями для доходов
1.2	Прогнозная	Будущие декларации: прогнозирование поздней подачи и неподачи деклараций	Кто подаст декларацию с опозданием, не подаст ее или самостоятельно заполнит ее после крайнего срока
1.3	Прогнозная	Будущие платежи: прогнозирование поздних платежей и случаев неуплаты	Кто произведет оплату с опозданием или не произведет оплату несмотря на наличие средств
1.4	Прогнозная	Будущие платежи: прогнозирование реакции в зависимости от типа интервенции	Например, SMS, звонок из колл-центра, заказное письмо или визит
1.5	Прогнозная	Будущая отчетность: прогнозирование занижения данных (вероятность)	Кто не будет соблюдать нормы, по видам налога (расчет вероятности)
1.6	Прогнозная	Будущая отчетность: прогнозирование размера корректировок (последствия)	Каким будет вероятный результат аудиторской проверки (расчет дискретного значения)
1.7	Прогнозная	Будущее обеспечение соблюдения норм: прогнозирование реакции налогоплательщиков	Кто, вероятно, будет возражать против корректировки сумм налога
1.8	Прогнозная	Сегменты и проблемы: прогнозирование агрегированного будущего поведения	Изменится ли ситуация с соблюдением норм и как это отразится на государственных доходах
<b>2</b>	<b>Адресная работа с клиентами</b>		
2.1	Прогнозная	Прогнозирование готовности соблюдать налоговые нормы (вероятность)	На основе истории соблюдения норм (налогоплательщик и связанные с ним структуры)
2.2	Прогнозная	Прогнозирование поступлений в результате работы с клиентом (последствия)	На основе предыдущей работы с клиентами и полученных поступлений, по каналам связи
<b>2.a</b>	<b>Следующие оптимальные шаги</b>		
2.a.1	Прогнозная	Варианты работы с клиентами	С указанием каналов, сроков и вероятных результатов
2.a.2	Прогнозная	Варианты административных мер	Определение последовательности действий, которые с наибольшей вероятностью приведут к разрешению вопроса (продолжение)

Таблица 3. Пример использования аналитики: активные меры предотвращения (окончание)

Необходимо?	Вид аналитики	Результат или функция	Примечания
<b>3</b>	<b>Встроенная аналитика</b>		
3.1	Гибридная	Электронные справки об отсутствии налоговой задолженности: оценка рисков, связанных с запросами	С использованием встроенной логики или профилей рисков из других систем
3.2	Прогнозная	Обслуживание налогоплательщиков: обучение чат-ботов решению налоговых вопросов	Перенаправление запросов сотрудникам и поиск информации в банке знаний
3.3	Прогнозная	Электронная подача деклараций; рекомендации по представлению данных перед подачей декларации	Анализ ключевых значений в сопоставимых группах налогоплательщиков
3.4	Прогнозная	Электронная подача деклараций; анализ рисков перед подачей декларации	Анализ декларируемых значений до подачи налогоплательщиком декларации
3.5	Гибридная	✓ Возвраты: оценка запросов и связанных с ними рисками	Содействие в автоматическом проведении оплаты или рекомендации мер контроля
3.6	Прогнозная	Финансовые механизмы: оценка запросов и связанных с ними рисками	С указанием того, следует ли предоставлять налогоплательщику финансовые услуги
<b>4</b>	<b>Аналитика в режиме реального времени</b>		
4.1	Прогнозная	Счета-фактуры в рамках взаимодействия B2B и B2C; анализ рисков, связанных с проведением операций	С учетом участников операций и характера счета-фактуры
4.2	Гибридная	Счета-фактуры в рамках взаимодействия G2B и G2C; анализ рисков, связанных с проведением операций	С учетом характера счета-фактуры и моделей поведения
<b>5</b>	<b>Сотрудничество с третьими сторонами</b>		
5.1	Гибридная	Электронное подтверждение; регистрация и справки	Подтверждение регистрации и соблюдения норм для третьих сторон

Источник: персонал МВФ.

Примечание. B2B – взаимодействие между предприятиями; B2C – взаимодействие между предприятием и потребителем; G2B – взаимодействие между государством и предприятием; G2C – взаимодействие между государственным органом; SMS – служба коротких сообщений.

Таблица 4. Пример использования аналитики: мониторинг и обнаружение

Необходимо? Вид аналитики		Результат или функция	Примечания
<b>1</b>	<b>Мониторинг основных вопросов соблюдения норм</b>		
1.1	✓ Описательная	Текущая ситуация с регистрацией: незарегистрированные налогоплательщики	На основе данных третьих сторон (например, реестра предприятий)
1.2	✓ Описательная	Текущая ситуация с регистрацией: требования регистрации по видам налога	С акцентом на НДС и другие важные пороговые значения
1.3	✓ Описательная	Текущая ситуация с регистрацией: налогоплательщики с недействующей регистрацией и активностью	На основе текущих данных третьих сторон (например, коммунальных служб)
1.4	✓ Описательная	Текущая ситуация с подачей деклараций: своевременная подача, подача с опозданием и неподача	Сводные и дезагрегированные данные (например, по налогоплательщикам или видам налога)
1.5	✓ Описательная	Текущая ситуация с платежами: своевременные, поздние платежи и случаи неуплаты	Сводные и дезагрегированные данные (например, по налогоплательщикам или видам налога)
1.6	✓ Гибридная	Текущая ситуация с представлением данных: случайный отбор кандидатов для проведения проверок	Для отслеживания уровней соблюдения норм с выявлением имеющихся проблем
1.7	✓ Описательная	Текущая ситуация с просроченной задолженностью: новая, zvyšиваемая и безнадежная задолженность	Сводные и дезагрегированные данные (например, по налогоплательщикам или видам налога)
1.8	✓ Описательная	Текущая ситуация с кредитованием: остаток, рост и неустраиваемые возвраты	Сводные и дезагрегированные данные (например, по налогоплательщикам или видам налога)
1.9	Описательная	Текущая ситуация с добровольным соблюдением норм: показатель повторных нарушений	Процентная доля налогоплательщиков, повторно не соблюдающих налоговые нормы
<b>2</b>	<b>Автоматическое обнаружение занижения данных</b>		
2.1	Гибридная	Автоматическое сопоставление данных: работодатели	Сопоставление указанных в декларациях значений с суммами, удержанными с работников
2.2	Гибридная	Автоматическое сопоставление данных: другие национальные и международные данные	Сопоставление внутренних и внешних данных (например, отчеты СВС)
<b>3</b>	<b>Разведочный анализ данных</b>		
3.1	✓ Описательная	Организация наборов данных	Определение элементов, размера и формы перед моделированием
3.2	✓ Описательная	Выявление взаимосвязей	Изучение характеристик и переменных на предмет возможных корреляций
3.3	✓ Описательная	Визуальное представление и обобщение	Использование графиков рассеивания, гистограмм, диаграмм и вспомогательных статистических данных
3.4	✓ Гибридная	Обнаружение резко отклоняющихся значений и аномалий	Для улучшения моделирования и выявления случаев несоблюдения норм (продолжение)

Таблица 4. Пример использования аналитики: мониторинг и обнаружение (окончание)

Необходимо?	Вид аналитики	Результат или функция	Примечания
<b>4</b>	<b>Аналитика профилей</b>		
✓	Описательная	Комплексный обзор: внутренние данные	Представление данных и показателей для презентации (например, по декларациям)
✓	Описательная	Комплексный обзор: внешние данные	Обеспечение согласованности с представлением внутренних данных
✓	Описательная	Комплексный обзор: дисперсия целевых показателей	Предварительный расчет отклонений от сопоставимых групп и контрольных наборов данных
✓	Описательная	Комплексный обзор: показатели налогоплательщиков	Из предварительно рассчитанного каталога показателей налогоплательщиков
✓	Описательная	Комплексный обзор: анализ рисков	Включает сводную информацию о рисках на основе проводимого вручную анализа
✓	Гибридная	Комплексный обзор: профили рисков	Интеграция результатов анализа рисков и предложение дальнейших действий
<b>5</b>	<b>Сотрудничество с третьими сторонами</b>		
✓	Гибридная	Специальный анализ для выявления рисков	Анализ данных третьих сторон без применения автоматизированных процессов

Источник: персонал МВФ.

Примечание: СВС — по отдельным странам; НДС — налог на добавленную стоимость.

Таблица 5. Пример использования аналитики: обеспечение соблюдения налоговых норм

Необходимо?	Вид аналитики	Результат или функция	Примечания
<b>1</b>	<b>Отбор кандидатов для проверок соблюдения норм</b>		
1.1	✓ Прогнозная	Текущая ситуация с регистрацией: прогнозирование приоритетов по принудительной регистрации	Категория дела: регистрация. Оценка последствий для доходов
1.2	Прогнозная	Текущая ситуация с регистрацией: прогнозирование случаев мошеннической регистрации	Категория дела: посещение. Приоритетность случаев регистрации новых плательщиков НДС
1.3	Прогнозная	Текущее отсутствие активности: прогнозирование вероятности фактического отсутствия активности	Категория дела: снятие с учета. В соответствии с определением отсутствия активности
1.4	Прогнозная	Отчетность за прошлые периоды: прогнозирование корректировок на основе несоответствия данных	Категория дела: проверка. Автоматическое обнаружение (несоответствие данных)
1.5	✓ Прогнозная	Отчетность за прошлые периоды: прогнозирование других приоритетов по занижению данных	Категория дела: аудиторская проверка. Оценка вероятности и последствий
1.6	✓ Прогнозная	Данные об оплате за прошлые периоды: прогнозирование вероятных результатов взыскания	Категория дела: взыскание. Приоритетность новой задолженности и задолженности с истекающим сроком погашения
1.7	✓ Прогнозная	Отчетность за прошлые периоды: прогнозирование мошеннических требований возвратов	Категория дела: возврат. Разбивка прогнозов по размеру требований
1.8	Гибридная	Формы для ведения дел	Наполнение статистикой инструментов, используемых ведущими дела сотрудниками
<b>1.a</b>	<b>Следующие оптимальные шаги</b>		
1.a.1	✓ Предписывающая	Параметры конфигурации дел: сфера деятельности	Предложение параметров (например, комплексная проверка или проверка по отдельным вопросам)
1.a.2	✓ Гибридная	Параметры конфигурации дел: выделение ресурсов	Рекомендации в отношении числа и способностей сотрудников, необходимых для ведения дела
1.a.3	Предписывающая	Параметры конфигурации дел: вероятные возражения	Предложения в отношении подходов, направленных на избежание возражений налогоплательщиков (апелляции)
<b>2</b>	<b>Эффективность отбора кандидатов для проверок</b>		
2.1	Гибридная	Оценка прогнозирующих возможностей	Использование матриц несоответствий и ROC-кривых
2.2	Описательная	Анализ результатов по соблюдению норм	Измерение изменений в поведении, обусловленных работой с налогоплательщиками
2.3	✓ Описательная	Анализ результатов по доходам	Измерение изменений в доходах, обусловленных работой с налогоплательщиками
2.4	✓ Гибридная	Возможности улучшения отбора кандидатов для проверок	Для включения в будущую логику отбора и аналитику (продолжение)

Таблица 5. Пример использования аналитики: обеспечение соблюдения налоговых норм (окончание)

Необходимо?	Вид аналитики	Результат или функция	Примечания
<b>3</b>	<b>Оценка и уведомления</b>		
3.1	Прогнозная	Оценка: прогнозирование значений оценок невыполнения обязательств	Составление оценок в рамках сопоставимых групп
3.2	Прогнозная	Оценка: прогнозирование возражений против оценок невыполнения обязательств	Включая определение пороговых значений для изменения поведения
3.3	Предписывающая	Уведомление: варианты уведомлений об оценке	Включая определение каналов и вариантов сообщений
3.4	Предписывающая	Уведомление: варианты уведомлений о взыскании средств	С учетом вариантов первого и последующих уведомлений
3.5	Предписывающая	Уведомление: варианты других уведомлений	С учетом желаемых результатов отправок уведомления
<b>4</b>	<b>Сотрудничество с третьими сторонами</b>		
4.1	Гибридная	Отбор налогоплательщиков для проверок: совместная проверка с таможенной службой	Использование специальных критериев и данных таможенной службы
4.2	Описательная	Сертификация для обмена данными: список налогоплательщиков для принудительного взыскания задолженности	Подтверждение включения в список налогоплательщиков с просроченной задолженностью, которую можно взыскать

Источник: персонал МВФ.

Примечание. ROC-кривая – кривая соотношения истинных и ложных характеристик; НДС – налог на добавленную стоимость.



## Приложение 1. Доступ к сопроводительному инструментарию

Инструментарий, разработанный в дополнение к этой справке, содержит настраиваемые примеры аналитических процедур для поддержания деятельности по трем направлениям, которые крайне важны для работы большинства налоговых администраций. Инструментарий, который можно скачать на Портале государственных доходов Департамента по бюджетным вопросам МВФ, включает проверенные на практике формы, настроенные на использование синтетических (искусственных) наборов данных. Оценив их работу, налоговые администрации, желающие продолжить изучение их возможностей, могут адаптировать конфигурацию форм для использования собственных внутренних данных. Со временем инструментарий будет обновляться и дополняться.

**Предполагается, что первоначально в него войдут следующие материалы:**

- **Форма МДР.** Файл в формате Microsoft Excel, содержащий готовый анализ МДР в виде наглядного графика рассеивания со сводной статистикой по квадрантам, предназначенный для использования при планировании обеспечения соблюдения налоговых норм крупными налогоплательщиками.
- **Форма для составления профилей крупных налогоплательщиков.** Для дальнейшего анализа относительно небольшого числа налогоплательщиков с высоким риском, выявленных в ходе анализа крупных налогоплательщиков с использованием МДР, разработан файл в формате Microsoft Excel, который можно использовать в качестве формы для составления профилей вручную.
- **Расширенный аналитический пакет для отбора налогоплательщиков для проведения проверок.** Для отбора кандидатов для проверок среди малых и средних налогоплательщиков разработан рабочий процесс, предполагающий использование платформы KNIME Analytics Platform и машинного обучения, а также алгоритм вычисления прогнозных возможностей.

**Для работы инструментария требуется следующее:**

- Последняя версия Microsoft Office и Excel<sup>5</sup>
- Последняя версия KNIME Analytics Platform<sup>6</sup>

### Инструкции для получения доступа к инструментарию

Шаг 1. Перейдите на Портал государственных доходов Департамента по бюджетным вопросам МВФ (<https://www.imf.org/revenueportal>).

Шаг 2. Перейдите в раздел портала “Analytical and Learning Resources” («Аналитические и образовательные ресурсы»).

Шаг 3. Найдите на странице «Essential Analytics for CRM» («Основные возможности аналитики для УКР»).

Шаг 4. Скачайте инструменты и инструкции.

На момент публикации прямой доступ к странице и разделу портала, где можно скачать инструментарий, доступен по следующей ссылке: <https://www.imf.org/en/Topics/fiscal-policies/Revenue-Portal/Analytical-and-Learning-Resources#analytical>

При возникновении трудностей просьба обращаться по адресу: [revenueportal@imf.org](mailto:revenueportal@imf.org).

<sup>5</sup> См. Microsoft Excel (<https://www.microsoft.com/en-us/microsoft-365/excel>).

<sup>6</sup> См. KNIME (<https://www.knime.com>).

## Приложение 2. Необходимые профильные знания

Для развития элементов аналитики, поддерживающих УКР, в дополнение к техническим навыкам в области статистики и науки о данных необходимы знания в четырех принципиально важных областях. Среди них методы выявления налогоплательщиков, представляющих интерес, использование статистических распределений для анализа налоговых данных и методы выявления резко отклоняющихся значений. В ближайшие годы в связи с развитием ИИ информация на эти темы станет более доступной, в частности благодаря использованию технологиями ИИ методов обработки естественного языка (NLP). Ниже представлена основная информация.

### Методы выявления налогоплательщиков, представляющих интерес

Большинство случаев использования аналитики для анализа рисков можно отнести к одной из пяти широких категорий, описывающих применяемые подходы. На практике эти категории представляют собой набор инструментов, доступных для использования специалистами по анализу данных, а также все чаще для интеграции в автоматизированные ИТ-сервисы, использующие встроенную аналитику. Независимо от конкретного применения, цель каждого подхода – выявить налогоплательщиков, представляющих интерес (как соблюдающих, так и потенциально не соблюдающих нормы), с помощью следующих методов:

- **Сопоставление.** Сопоставление данных и объединение наборов данных, включая качественную аналитику. Сопоставление данных широко используется для решения некоторых видов задач, таких как выявление незарегистрированных лиц и лиц, не представляющих налоговые декларации. Оно может предусматривать сопоставление данных о предприятиях (сравнение базы данных лицензий на ведение бизнеса с налоговой базой данных или базы данных телефонного справочника с налоговой базой данных). Сопоставление может быть как «жестким» (с использованием идентификационных номеров), так и «мягким» (с использованием имен). Результаты сопоставления данных часто требуют определения приоритетов (например, путем оценки сумм налога).
- **Измерение.** Определение и расчет показателей и значений на основе экспертных оценок и правил. Другой подход, часто используемый налоговыми администрациями, предполагает выявление значимых атрибутов в наборе данных и измерение их значений. Обычно такие атрибуты определяются персоналом, имеющим значительный опыт в этой области, на основе их значимости для комплаенс-контроля. Например, суммы возвратов, превышающие определенный уровень и отклоняющиеся от типичных значений, часто являются сигналом для проведения комплаенс-проверки. Результаты таких измерений нередко служат основой для работы «механизмов выполнения правил».
- **Интеллектуальный анализ.** Выявление значимых кластеров или сегментов с помощью различных методов. Обычно они применяются к большим и сложным наборам данных для выявления новых групп, которые могут представлять интерес, и лиц, которые выделяются из общей группы. К таким методам относятся метод k-ближайших соседей (например, для обнаружения резко отклоняющихся значений в группе или предупреждения налогоплательщиков о том, что значения, указанные в их декларациях, отклоняются от контрольных ориентиров по отрасли) и анализ текстовых массивов (например, для анализа большого количества электронных писем и выявления тех из них, которые имеют значение для механизмов налогового планирования).
- **Моделирование.** Построение прогнозных моделей для классификации вероятности и последствий потенциального несоблюдения норм. Результаты прошлых проверок анализируются с помощью одного

или нескольких алгоритмов, часто с использованием машинного обучения, для прогнозирования статуса налогоплательщика (соблюдает или не соблюдает нормы), а также вероятности того, что прогнозы верны. По мере увеличения объема и сложности данных использование таких подходов становится все более важным. При наличии достаточного объема данных эти методы часто более эффективны по сравнению с использованием «правил оценки рисков».

- **Установление взаимосвязей.** Использование специализированного программного обеспечения для анализа сетей с целью установления связей и отслеживания цепочек данных. В рамках сети такие методы помогают выявлять значимые узлы. В налоговом администрировании существует множество способов их применения. В качестве примеров можно привести отслеживание данных об операциях по налогу на добавленную стоимость, проводимых между организациями, для выявления случаев мошенничества с отсутствующим торговцем, а также использование анализа социальных сетей (не путать с онлайн-платформами) для обнаружения мошеннических организаций, скрывающих свою деятельность.

### Полезные статистические распределения для анализа налоговых данных

**Для получения полезных результатов анализа важное значение имеют правильный выбор и использование распределений.** Это особенно важно при использовании методов индуктивного и прогнозного моделирования. Хотя выбор наиболее точного распределения для моделирования налоговых рисков всегда зависит от конкретных данных и постановки исследовательских вопросов, существует ряд общепринятых концепций и методов их использования:

- **Логнормальное распределение.** Обычно используется для моделирования данных, полученных в результате накопления множества событий с малым риском. Хотя для использования логнормального распределения данные не могут быть отрицательными, эту проблему можно решить простым смещением «нулевой точки» до наибольшего отрицательного значения. Умножение переменных обычно приводит к возникновению логнормального распределения. А риск – это по определению вероятность, умноженная на последствия.
- **Распределение Парето.** Обычно используется для моделирования рисков, распределение которых подчиняется степенному закону, что означает, что небольшому количеству событий соответствует высокая доля риска. В связи с этим это распределение часто используется для моделирования таких данных, как уклонение от уплаты налогов и мошенничество, а также размер корректировок. Оно также может использоваться для моделирования распределения доходов или богатства, либо для избежания налогообложения, особенно крупными налогоплательщиками.
- **Распределение Вейбулла.** Часто используется для моделирования данных о времени до наступления события. Например, для моделирования времени до начала несоблюдения норм налогоплательщиком или времени, затрачиваемого на обнаружение и урегулирование случая уклонения от уплаты налогов. Оба подхода могут быть полезны при изучении потенциальных результатов комплаенс-стратегий и планов, а также использования ресурсов.
- **Бета-распределение.** Гибкое распределение вероятностей, которое позволяет моделировать данные, ограниченные двумя значениями, например долю налогоплательщиков, не соблюдающих налоговые нормы, или долю налоговых поступлений, теряемых из-за несоблюдения налоговых норм.
- **Мультиномиальное распределение.** Может использоваться для моделирования категориальных данных, таких как статус налогоплательщиков, в том числе «соблюдает нормы», «частично соблюдает нормы» (ниже порогового уровня) или «не соблюдает нормы» (выше порогового уровня). После этого можно выполнить оценку генеральной совокупности на основе выборок с использованием байесовской оценки решения или оценки максимального правдоподобия.

- **Распределение Пуассона.** Используется для оценки вероятности наличия в генеральной совокупности определенного числа налогоплательщиков, не соблюдающих налоговые нормы. Например, поскольку это распределение может моделировать количество событий, происходящих за определенный период времени, его можно использовать для прогнозирования числа аудиторских проверок или числа налогоплательщиков, не соблюдающих нормы, за определенный период.
- **Экспоненциальное распределение.** Еще одно распределение для моделирования данных о времени до наступления события, например времени до начала несоблюдения налогоплательщиком норм («коэффициент риска» несоблюдения норм с течением времени), отличающееся от распределения Вейбулла использованием вероятностей.
- **Распределение Гаусса или нормальное распределение.** Это распределение, вероятно, наиболее знакомо большинству сотрудников налоговых служб из базовых курсов статистики. Однако нормальное распределение не показывает распределение большинства атрибутов данных налогоплательщиков, поскольку они отражают базовые экономические факторы, которые, в свою очередь, обычно имеют распределение Парето.

### Выявление резко отклоняющихся значений с помощью разведочного анализа данных

Одним из основных элементов аналитики для УКР является концепция разведочного анализа данных (РАД) (которая также упоминается в таблице 4). Цель разведочного анализа данных – выявить, понять и обобщить основные характеристики наборов данных, часто с использованием визуальных методов. Он является важной предпосылкой для использования статистики в любых целях УКР. С практической точки зрения, прежде чем делать выводы на основе конкретных данных, эти данные необходимо хорошо понять. Для этого в рамках разведочного анализа данных используются различные методы, позволяющие обобщить характеристики данных и изучить их пригодность. Такие методы часто включают выявление закономерностей путем корреляционного анализа, снижение размерности, кластеризацию данных и другие подходы.

**В рамках разведочного анализа данных важное значение имеет обнаружение аномалий и резко отклоняющихся значений.** Их наличие может влиять на решения о пригодности данных (и их потенциальной корректировке), а также указывать на случаи несоблюдения налоговых норм. К методам их обнаружения, которые важны для многих аспектов УКР, относятся следующие:

- **Z-оценка.** Этот метод позволяет определить расстояние (в стандартных отклонениях) между средним значением набора данных и каждой точкой данных. Точки данных, удаленные от среднего значения более чем на определенное число стандартных отклонений, считаются резко отклоняющимися значениями.
- **Межквартильный диапазон.** Этот метод позволяет рассчитать разницу между 75-м и 25-м percentилями набора данных. Точки данных, которые выходят за пределы межквартильного диапазона ( $Q3 + 1,5$  межквартильного диапазона) и ( $Q1 - 1,5$  межквартильного диапазона), считаются резко отклоняющимися значениями. Можно использовать и другие диапазоны, при этом важно анализировать влияние изменения параметров.
- **Расстояние Махаланобиса.** Этот метод позволяет определить расстояние между средним значением набора данных и каждой точкой данных с учетом ковариации набора данных. Точки данных, удаленные от среднего значения более чем на определенное пороговое значение, считаются резко отклоняющимися значениями.
- **Локальный уровень выброса.** Этот метод основан на концепции плотности и определяет резко отклоняющиеся значения путем сравнения плотности точки данных с плотностью ее соседей. Точки данных, плотность которых значительно ниже, считаются резко отклоняющимися значениями.

- **Эллиптическая огибающая.** Это метод, основанный на многомерном нормальном распределении. Он предполагает, что данные имеют нормальное распределение, и позволяет выявить точки данных, удаленные от центра. Этот метод требует нормализации данных.
- **Методы кластеризации.** Точки данных, которые слабо представлены в каком-либо кластере или находятся далеко от центроида какого-либо кластера, считаются резко отклоняющимися значениями. Существует несколько методов кластеризации, включая кластеризацию k-средних, иерархическую кластеризацию, плотностный алгоритм кластеризации пространственных данных с присутствием шума и алгоритм «изолирующий лес».

## Использование ИИ для обработки естественного языка

**Алгоритмы машинного обучения, основанные на десятилетиях исследований, позволяют компьютерам понимать и обрабатывать команды на естественном языке.** Эта концепция обычно рассматривается как уникальное направление ИИ, называемое NLP. В своей современной форме технология NLP использует алгоритмы, разработанные для «обучения с подкреплением» на крупных нейронных сетях, которые обучаются на больших объемах текстовых данных (большие языковые модели). Возможности NLP имеют большой потенциал в сфере налогового администрирования.

**В сфере обслуживания налогоплательщиков внедрение этой технологии идет быстрыми темпами.** Наиболее заметно это проявляется в использовании чат-ботов на базе алгоритмов NLP. Эти боты оказывают автоматизированную и персонализированную помощь, отвечая на вопросы и предоставляя рекомендации по вопросам, связанным с налогообложением. Аналогичные возможности стремительно внедряются и в системы колл-центров.

**Для поддержки деятельности в области УКУ существуют практические способы применения этой технологии, которые уже используются в ограниченном объеме (например, для формирования запросов к базам данных и выполнения других задач программирования для целей аналитики).** По мере своего развития NLP и инструменты, в которых используется эта технология, вероятно, будут оказывать преобразующее воздействие на деятельность в области рисков и аналитики благодаря следующим возможностям:

- **Понимание разных областей знаний.** Технология NLP способна понимать вопросы, задаваемые на естественном языке, которые охватывают широкие области специализированных знаний (налогообложение, право и бухгалтерский учет).
- **Динамическое создание кода.** При условии, что вопросы предполагают получение выводов или прогнозов на основе данных, технология NLP способна формировать запросы к данным и сценарии для аналитики.
- **Применение логики к большим наборам данных.** Используя генерируемые запросы, инструменты с поддержкой NLP могут самостоятельно находить ответы на вопросы в данных.
- **Интерпретация результатов.** Помимо извлечения нужных данных, система способна интерпретировать результаты в контексте исходного вопроса и генерировать соответствующие ответы с использованием NLP.

**Хотя в настоящее время использование технологии NLP в сочетании с ботами в налоговых ведомствах сводится преимущественно к вопросам обслуживания, их адаптация для поддержки специалистов в области УКУ вполне вероятна.** На практике боты, работающие на базе NLP, в дальнейшем могут обеспечить получение точных и быстрых ответов на прямые вопросы, такие как:

- Каковы основные риски несоблюдения налоговых норм в моей юрисдикции?
- Насколько эффективны мои стратегии по обеспечению соблюдения налоговых норм?
- Какие налогоплательщики наиболее склонны к занижению данных в отчетности?

**Использование NLP не обходится без издержек, так как требует выбора используемых ботов и инструментов, их обучения и тонкой настройки, корректного формулирования вопросов и обеспечения качества ответов.** Хотя в настоящее время эти вопросы представляют значительные трудности, в использовании ИИ наблюдается стремительный прогресс, а инструменты с поддержкой ИИ быстро становятся более доступными. Этому прогрессу способствует ряд инициатив, которые уже реализуются в целях более широкого использования ИИ, в частности в налоговом администрировании.

## Приложение 3. Решение ключевых задач: качество данных и кадровый потенциал

В разных странах мира особую трудность для многих налоговых администраций представляют две сохраняющиеся проблемы: качество данных и кадровый потенциал. Хотя для этих проблем, как правило, не существует простых решений, в последние годы реализуется множество серьезных инициатив по улучшению аналитических возможностей в области УКР. В результате накопленного опыта можно выделить несколько ключевых концепций и методов, которые известны в разных регионах мира как более эффективные по сравнению с другими.

### Качество данных

Вопросы качества данных, которые отражают уровень развитости институтов, следует решать с помощью четких определений и стратегий постепенного повышения эффективности. Все налоговые ведомства являются организациями, деятельность которых основана на информации. Данные используются повсеместно: от основной операционной деятельности по налогообложению до стратегии и планирования. С концептуальной точки зрения полученные данные должны быть результатом реализации четко определенных методов работы, которые, в свою очередь, подлежат оценке и контролю качества. В тех случаях, когда процессы четко не определены и не реализуются в полной мере, качество данных часто бывает проблематичным. Независимо от причин, например низкой эффективности ИТ-систем, низкой квалификации персонала и слабой подготовки кадров, неэффективное управление процессами обычно является следствием институциональных недостатков.

Качество данных имеет важное значение не только непосредственно в контексте налогового администрирования, но также для государственного сектора и предприятий. Особенно это касается тех областей, где для операционных целей (обработка платежей) и ключевых аспектов анализа рисков или автоматизации процессов на основе рисков (автоматическое обнаружение случаев занижения данных в отчетности) используются данные третьих сторон.

В области аналитики для УКР «качество данных» можно определить как степень пригодности данных для поддержки анализа рисков. Соответственно, качество данных можно измерить. Во вставке 3.1 приложения в качестве иллюстрации представлены девять построенных на принципах элементов основы качества данных. В ведомствах с развитой практикой управления данными наборы данных формально регистрируются и оцениваются в соответствии с описанными элементами. После этого определяется общий рейтинг качества (например, низкий, средний, высокий), а также отдельные рейтинги, описывающие полезность данного набора данных для анализа рисков. Несмотря на значительные инвестиции в цифровые технологии, на сегодняшний день во многих администрациях эта практика пока не применяется.

Поскольку качество данных зависит от многих аспектов управления, наиболее успешными, как правило, являются стратегии, которые в первую очередь уделяют внимание основным принципам. Несмотря на важность технических аспектов вопросов, связанных с качеством данных, обязанность по его улучшению лежит на всей организации, а не исключительно на отделе ИТ. Соответственно, как правило, требуется участие технических специалистов и лиц, ответственных за соответствующие процессы, системы и данные. В рамках официальной стратегии, одобренной руководством, могут поэтапно приниматься следующие меры, приведенные в примерной последовательности:

- **Внедрение основных методов управления данными.** К ним относятся инвентаризация существующих информационных активов, их классификация по уровням важности и конфиденциальности,



**ВСТАВКА 3.1 ПРИЛОЖЕНИЯ. Элементы основы качества данных**

Элемент основы	Качество данных отражает степень, в которой...
Достоверность	данные с точностью отражают фактические значения.
Полнота	необходимые данные действительно представлены в наборе данных.
Соблюдение норм	в отношении набора данных действуют правила, меры контроля и система сквозной регистрации операций.
Согласованность	значения набора данных согласуются с другими имеющимися данными.
Детализация	уровень детализации элементов набора данных достаточен.
Актуальность	данные за период времени, охватываемый набором данных, актуальны.
Отслеживаемость	сохраняется история изменений данных в наборе данных.
Уникальность	данные дублируются в наборе данных или их копии хранятся отдельно.
Допустимость	данные представлены в надлежащем формате и с учетом установленных ограничений.

Источник: персонал МВФ.

определение каталога данных и словарей данных, а также разработка основных стратегий в отношении жизненного цикла данных.

- **Определение и оценка базового качества основных наборов данных.** К основным наборам данных относятся либо важная справочная информация (виды налогов, налоговые периоды, налоговые формы), либо данные, необходимые для ведения основной деятельности (налоговый реестр, исчисление налога, счета, возвраты, платежи).
- **Определение коренных причин основных проблем, связанных с качеством данных.** При этом важно отделять проблемы в области ИТ от проблем, связанных с базовыми рабочими процессами и тем, как они используются персоналом и налогоплательщиками (последовательно или нет).
- **Реализация целевого плана по повышению эффективности.** План, ориентированный только на основные наборы данных, должен устанавливать приоритетность интервенций (например, пересмотр порядка использования ИТ-систем), направленных на повышение качества текущих данных, а не данных за прошлые периоды.
- **Включение в оценку неосновных наборов данных.** Сюда следует отнести все остальные наборы данных, для каждого из которых устанавливается приоритетность проведения анализа, при этом основное внимание следует уделять данным, полученным от третьих сторон, данным о работе с налогоплательщиками, корреспонденции и объемной электронной отчетности (например, счета-фактуры).
- **Реализация плана по повышению эффективности учреждения.** Целесообразные меры, состоящие из компонентов, которые могут представлять собой значительные реформы, могут предусматривать внедрение систем обеспечения качества, новые процедуры, подготовку кадров и приобретение инструментов.



- **Работа с данными внешних партнеров и третьих сторон.** Учитывая, что организации, предоставляющие данные, также могут сталкиваться с трудностями, формальный процесс сертификации источников данных, определений данных и фактически предоставленных данных должен предусматривать соответствующие меры контроля качества.

**На этапе внедрения улучшений специалистам следует проводить четкую грань между проверенными качественными наборами данных и остальными данными, а также опираться на человеческий интеллект.** В юрисдикциях с ограниченными или некачественными данными следует применять избирательный подход к их использованию во избежание негативных последствий. Пока стратегии в отношении качества данных не начнут действовать, для анализа налоговых деклараций и платежей специалистам следует опираться на базовую описательную статистику. В отношении случаев занижения данных вместо более сложных методов прогнозирования можно использовать экспертные правила. Базовое сопоставление данных также может быть весьма эффективно. По мере появления более качественных наборов данных следует изучить весь спектр аналитических возможностей и рассмотреть вопрос об их внедрении.

## Кадровый потенциал

**Поскольку аналитика носит междисциплинарный характер, недостатки потенциала лучше всего устранять путем сочетания набора сотрудников, подготовки кадров и консолидации ресурсов.** В идеальной ситуации поддержку УКР должны оказывать подготовленные специалисты по обработке и анализу данных. Как показано на рис. 2, наука о данных опирается на опыт из трех областей знаний: информатики, бизнеса и статистики. Для эффективной поддержки УКР специалисту по обработке и анализу данных необходимы профессиональные навыки в двух областях с математическим уклоном (информатика и статистика), а также знания в области налогового администрирования. Хотя уже имеющиеся знания новых сотрудников в области информатики и статистики могут быть применимы на новой должности, освоение технических аспектов налоговых вопросов, операций налогового администрирования и УКР может занять значительное время.

**В то время как более крупные налоговые службы все чаще нанимают непосредственно специалистов по обработке и анализу данных и проводят их переподготовку по вопросам налогообложения, другие ведомства, как правило, обучают сотрудников необходимым навыкам самостоятельно.** Организации как государственного, так и частного сектора в целом, как правило, рассматривают науку о данных как область, имеющую стратегическое значение. Высокая ценность, присваиваемая требуемым навыкам, создала конкуренцию практически во всех регионах мира, в результате чего оплата труда соответствующих специалистов возросла до уровней, которые многие налоговые службы не могут себе позволить. Несмотря на сложные условия, эти трудности с определенным успехом удается преодолевать с помощью перечисленных ниже креативных стратегий найма и нетрадиционных стимулов:

- **Маркетинг уникального масштаба и объема управляемых данных.** Донесение информации о том, что во многих юрисдикциях налоговые ведомства являются одними из крупнейших ИТ-операторов в стране и располагают данными, которые могут представлять уникальную интеллектуальную ценность, особенно в контексте расследовательской и правоприменительной деятельности.
- **Маркетинг присутствия специалистов по обработке и анализу данных на важных заседаниях комитетов.** Демонстрируя ценность требуемых навыков, включение аналитиков в работу фронт-офиса (комплаенс-комитетов) подтверждает, что их роль не будет сводиться к поддержке бэк-офиса.
- **Предложение сотрудникам, имеющим образование в сфере ИТ, но работающим на должностях, не связанных с ИТ, перевестись на новую должность.** Выявление сотрудников, получивших образование или прошедших подготовку по получению необходимых навыков, занятых на нетехнических должностях, и обращение к ним с предложением перейти в группу аналитики по УКР.

- **Набор сотрудников из государственных учреждений, университетов и исследовательских центров.** Поиск специалистов по обработке и анализу данных, уже работающих в государственном секторе, и профессионалов, занятых в сфере высшего образования (например, исследователей, работающих в университетах), которым может быть интересна практическая деятельность.

**Независимо от используемого подхода к набору сотрудников, для содействия развитию потенциала необходимо определение целевого профиля сотрудников и анализ ситуации.** Если качество имеющихся данных недостаточно высоко для использования расширенной аналитики или практика УКР еще недостаточно развита, целесообразно начать с малого. В некоторых администрациях выделение любых ресурсов на цели аналитики в области УКР уже является серьезным шагом вперед и позволяет сократить зависимость от ИТ-персонала в работе с данными и информацией. В таких случаях требуется лишь базовая подготовка кадров. В других случаях необходимо формирование целой группы специалистов по обработке и анализу данных. С целью содействия в проработке сценариев во вставке 3.2 приложения представлен ряд профилей сотрудников и возможные шаги по повышению их потенциала в области аналитики.

**В развитии кадрового потенциала также могут помочь организационные стратегии: одним из перспективных вариантов является консолидация подразделений.** Во всех налоговых ведомствах анализ данных проводится по разным направлениям, особенно в головных офисах. Такая ситуация типична и требует, чтобы сотрудники каждого подразделения по отдельности развивали навыки по работе с данными и управлению их использованием. В последние годы в качестве одного из методов развития потенциала обсуждается идея объединения сотрудников, ответственных за 1) управление данными, 2) аналитику и 3) исследовательские функции, в единую группу. Такой подход потенциально может иметь значительные преимущества для целенаправленного развития потенциала, и в некоторых администрациях для расширения деятельности в области аналитики потребность в консолидации может возрасти.

### **ВСТАВКА 3.2 ПРИЛОЖЕНИЯ. Целевые профили сотрудников и возможности развития потенциала**

**Начинающий аналитик.** Опытный пользователь или эксперт в области Microsoft Excel, способный выполнять сложный анализ в Excel или аналогичных инструментах с использованием широкого набора формул, макросов и сценариев. Может обладать или не обладать экспертными знаниями в области бизнеса. Не способен формировать запросы к базам данных или анализировать большие наборы данных без поддержки ИТ-персонала. Сотрудник с таким профилем может:

- пройти тренинги для новых сотрудников и аудиторов для расширения знаний о налогах;
- пройти базовый тренинг по SQL для развития навыков формирования запросов к базам данных;
- запросить прямой доступ к базам данных для ускорения разведочного анализа;
- перенять опыт у администраторов баз данных для овладения навыками работы с реестром данных и моделями.

**Аналитик среднего уровня.** Специалист по работе с данными, обладающий уверенными навыками формирования запросов к базам данных и базовыми знаниями в области расширенной аналитики. Имеет полное и исчерпывающее представление обо всех основных администрируемых налогах, операционной деятельности налоговой администрации и управляемых наборах данных. Способен самостоятельно составлять статистику и отчеты, а также проводить анализ рисков без помощи ИТ-персонала. Сотрудник с таким профилем может:

- пройти углубленный тренинг по SQL для достижения полного профессионального уровня владения соответствующими навыками;
- пройти обучение по расширенной аналитике для улучшения навыков использования методов прогнозирования;
- пройти обучение по визуализации данных для улучшения навыков представления результатов;
- участвовать в мероприятиях по обмену знаниями с другими налоговыми администрациями;
- сотрудничать с другими ведомствами для расширения знаний в смежных областях;
- вести документацию о данных, инструментах и методах, используемых для оказания содействия другим сотрудникам.

**Квалифицированный специалист по обработке и анализу данных.** Эксперт, в полной мере освоивший информатику, налогообложение и статистику, понимающий математические расчеты и логику применяемых методов, а также знающий, в каких случаях их следует применять. Специализируется на прогнозной аналитике, проводит и тестирует анализ рисков и демонстрирует его эффективность с использованием больших объемов данных. Сотрудник с таким профилем может:

- пройти курсы повышения квалификации по налогообложению, чтобы использовать полученные знания при разработке новых методов анализа;
- подписаться на журналы, чтобы быть в курсе новых концепций и методов;
- приобретать технологии и инструменты для внедрения новых, специализированных возможностей.

Источник: персонал МВФ.

Примечание. SQL – язык структурированных запросов.

## Приложение 4. Отдельная тема. Комплаенс-планирование

В настоящей справке «комплаенс-планирование» представляет собой общий термин, используемый для описания разработки стратегий управления рисками и распределения ресурсов. Под директивным руководством исполнительного руководства и на основе анализа рисков и аналитических данных специалисты, оказывающие содействие комплаенс-планированию, определяют, каким рискам следует уделить внимание, какие ресурсы необходимо задействовать, как их следует применять и по каким направлениям могут понадобиться дополнительные действия. Результатами их работы могут быть планы по обеспечению соблюдения норм или увеличение объема инвестиций в проводимые стратегии и кампании. Эта деятельность часто осуществляется через официальные комитеты.

В качестве исходных элементов комплаенс-планирования применяются специализированные макро-стратегические концепции: среди них ключевую роль играет применение МДР в отношении крупного бизнеса. В этом разделе представлена современная концепция МДР. Ее логика основана на понимании того, что разным налогоплательщикам соответствуют разные уровни и типы рисков и что ресурсы должны распределяться с учетом этого. При логической форме матрицы рисков в модели используется аналитика, предполагающая выявление налогоплательщиков с более высокой вероятностью возникновения проблем (часто на основе низких эффективных ставок налога) и концентрацию на них соответствующих усилий по обеспечению соблюдения норм, в то время как в отношении налогоплательщиков с более низкой вероятностью возникновения проблем применяются упрощенные варианты мер по улучшению соблюдения норм (информирование, подтверждение соблюдения норм, консультации, другие улучшенные услуги). Как показано на рис 4.1 приложения, эти действия сгруппированы по четырем квадрантам, в описании которых представлен общий подход к обеспечению соблюдения норм (например, «предотвращение в реальном времени»).

Рисунок 4.1 приложения. Модель дифференциации рисков



**Использование МДР, которое все чаще встречается в разных странах мира, отражает важное развитие практики управления рисками, происходящее на протяжении многих десятилетий.** Сама по себе дифференциация рисков не является новой концепцией. Так, в 1980-х годах Австралийское налоговое управление ввело систему ранжирования SCORE, которая позволяла выявлять отклонения по отраслям и видам деятельности. Аналогичным образом, в 1990-х годах Налоговое управление США применило дифференциацию, разработав систему дискриминантной индексной функции для классификации деклараций о доходах для целей аудита по категориям высокого, среднего и низкого риска. Эти подходы были и во многих странах остаются основой для управления рисками. С появлением новых методов УКР в середине 2000-х годов концепции дифференциации рисков получили развитие по следующим направлениям

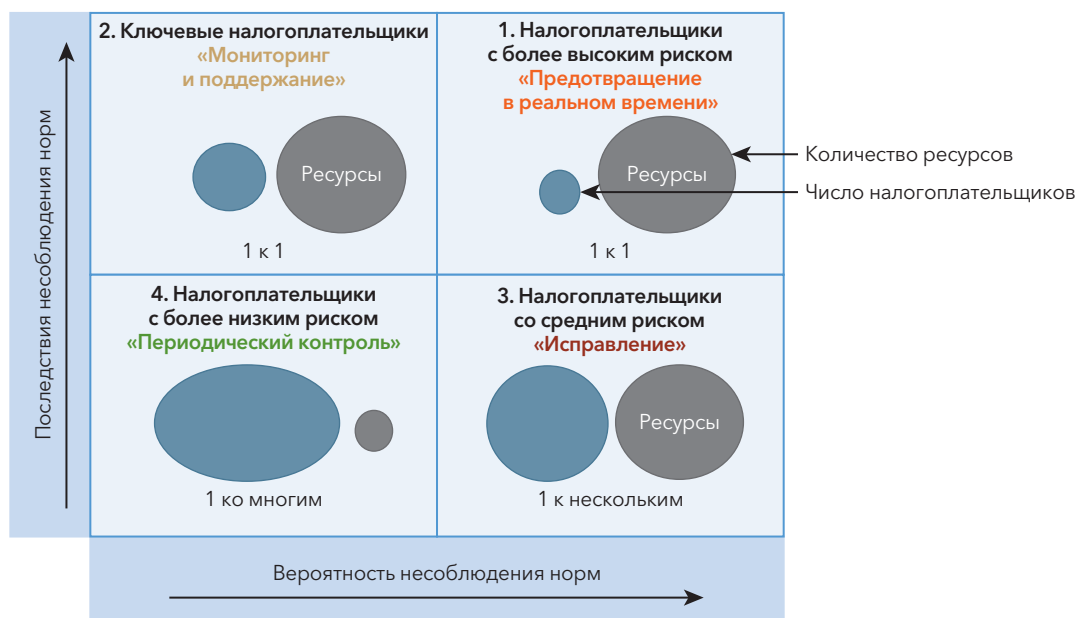
**Интеграция вероятности и последствий.** Позволяет учитывать оба аспекта риска, которые необходимо рассматривать в совокупности для проведения эффективных интервенций (но которые отсутствуют в анализе рисков, проводимом сегодня во многих налоговых администрациях).

**Распознавание моделей общего поведения.** Позволяет визуально представить результаты в виде графика рассеяния (часто наглядно демонстрирующего интересные закономерности) и сводной статистики по четырем квадрантам с выделением уникальных проблем в каждом из них.

**Поддержка перспективного комплаенс-планирования.** Благодаря своей способности выявлять закономерности и поведенческие различия путем визуального представления данных (через четыре квадранта) МДР особенно полезна для понимания и планирования мер в отношении разных групп налогоплательщиков.

**При рассмотрении концепции МДР с точки зрения ресурсного планирования вся ее полезность становится очевидна.** В рамках анализа МДР распределение налогоплательщиков, показанное синими сферами на рис. 4.2 приложения, должно привести к тому, что в квадранте 1 (налогоплательщики с более высоким риском) окажется мало налогоплательщиков, в квадрантах 2 и 3 (ключевые налогоплательщики и налогоплательщики со средним риском) – больше, а в квадранте 4 (налогоплательщики с более низким риском) – большинство. С учетом разных проблем, вероятностей и последствий наступления рисков в каждом квадранте эффективное

**Рисунок 4.2 приложения. Распределение ресурсов по квадрантам МДР**



Источник: персонал МВФ.

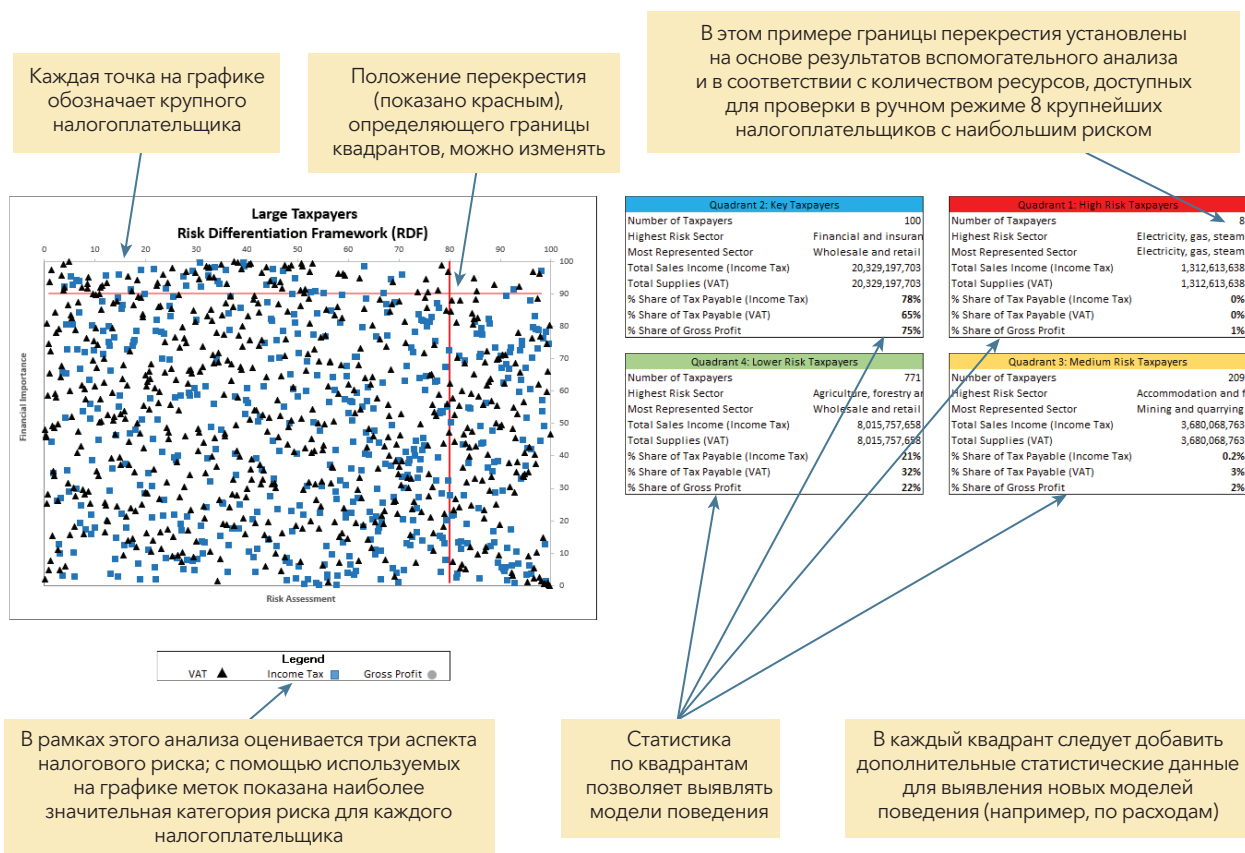
Примечание. МДР – модель дифференциации рисков.

использование ресурсов (прежде всего персонала) часто предполагает распределение, показанное серыми сферами. Для квадранта 1 целесообразен подход «один к одному», при котором управление рисками осуществляется отдельно по каждому налогоплательщику. Противоположная ситуация наблюдается в квадранте 4, где анализ проводится в отношении группы в целом.

При использовании МДР важно учитывать, что она предназначена для использования в качестве модели перспективного планирования и не обязательно как инструмента отбора кандидатов для проведения проверок. По завершении первоначальной аналитической работы по проведению анализа МДР (в соответствии с формой из инструментария) аналитикам следует уделить время на следующее: 1) интерпретацию выявленных визуальных и статистических закономерностей и 2) рассмотрение вариантов дальнейших оптимальных действий в рамках всестороннего процесса УКР. В некоторых случаях может быть целесообразно проведение аудиторских проверок. Однако существует целый ряд инструментов, которые могут быть более эффективны, особенно в отношении крупных налогоплательщиков (например, программы для ключевых клиентов, предварительные соглашения о ценообразовании, разъяснения, предложения по внесению изменений в законодательство). Выбранные меры необходимо включать в комплаенс-план.

Поскольку МДР – это прежде всего инструмент планирования, предназначенный для распределения ресурсов в зависимости от риска, для его эффективного использования важны суждения аналитиков. Конфигурацию модели можно и нужно настраивать с учетом применения различных показателей для рассмотрения разных видов налогов по отдельности и в рамках комплексного анализа. На рис. 4.3 приложения представлен пример использования формы из инструментария, разработанного в дополнение к настоящей справке.

Рисунок 4.3 приложения. Пример результатов применения формы МДР из инструментария



Источник: персонал МВФ.

Примечание. МДР – модель дифференциации рисков; НДС – налог на добавленную стоимость.

## Приложение 5. Отдельная тема. Профилирование налогоплательщиков

Для поддержки работы специалистов по анализу рисков все чаще требуются автоматическое профилирование и цифровые инструменты, позволяющие обрабатывать большие объемы данных и управлять анализом рисков. В отношении крупных налогоплательщиков идеальный для большинства администраций набор возможностей позволил бы специалисту выполнять следующие задачи: 1) изучение большого сегмента с помощью анализа МДР, 2) изучение подробного профиля каждого налогоплательщика на графике рассеяния, полученного на основе МДР, в интерактивном режиме и 3) отбор небольшого числа налогоплательщиков для дальнейшего ручного анализа рисков (например, представленных в квадранте 1 МДР, то есть налогоплательщиков с более высоким риском). С помощью цифровых инструментов результаты изучения случаев, выбранных для ручного анализа, следует вносить в каталог для дальнейшего анализа и улучшения работы системы с течением времени. Как показано выше на рис. 3 в основном тексте и далее во вставке 5.1 приложения, автоматическое профилирование опирается на слои инфраструктуры данных для получения «комплексного обзора» налогоплательщика.

Среди комплексных обзоров, которые в совокупности составляют профиль налогоплательщика, как правило, недостаточно развиты или отсутствуют два вида обзоров, связанных с рисками («анализ рисков» и «профиль рисков»). Этот факт неизменно отмечается как одно из ограничений в работе налоговых администраций по всему миру, при этом причины варьируются от низкого уровня использования автоматизации в целом до трудностей в разработке требований к необходимым ИТ-средствам и цифровым инструментам. В ведомствах, где автоматизация управления рисками предусмотрена, ее применение часто ограничивается планированием использования ресурсов и очень простой логикой отбора налогоплательщиков для проверок. Таким образом, несмотря на потенциальное наличие аналитических возможностей, необходимых для формирования комплексного обзора, и возможность их автоматизации, цифровые инструменты, с помощью которых специалисты могли бы активно управлять профилем налогоплательщика, зачастую отсутствуют полностью.

Целесообразным первым шагом на пути к улучшению цифровых инструментов для профилирования является принятие формы анализа рисков крупных налогоплательщиков, в чем может помочь модель BISEP (компания, отрасль, социологические, экономические, психологические факторы). Крупные налогоплательщики имеют множество характеристик, требующих анализа, выходящего далеко за рамки аналитики. Хотя существуют разные методы их оценки, часто используется концепция BISEP. Как показано на рис. 5.1 приложения, BISEP – это акроним, охватывающий пять факторов, влияющих на поведение налогоплательщика (начиная с «компании» и далее по часовой стрелке). Подробный анализ крупного налогоплательщика сам по себе является исследовательским проектом, и BISEP активно используется для этой цели как наиболее развитыми, так и развивающимися администрациями. В некоторых случаях модель BISEP была расширена за счет включения дополнительных факторов, таких как комплаенс-системы, технологии и данные.



### **ВСТАВКА 5.1 ПРИЛОЖЕНИЯ. Авторитетные комплексные обзоры налогоплательщиков**

В настоящее время многие налоговые службы вкладывают значительные средства в развитие инфраструктуры данных. По мере роста объема управляемых данных все чаще требуются специализированные технологии, инструменты и методы моделирования данных. Эта динамика особенно выражена в ряде областей, включая концепцию «комплексного обзора». В связи с частой необходимостью быстрого получения полного набора данных о налогоплательщике или поиска информации по нему в хранилище данных часто размещается набор вспомогательных моделей данных. Хотя технические подходы различаются, концептуально дискретные модели часто разрабатываются специально для быстрого получения отдельных авторитетных представлений данных:

- **Комплексный обзор: внутренние данные.** Все данные, созданные или полученные самим ведомством. Наборы данных, представленные в структурированном формате (то есть не являющиеся необработанными данными), охватывают все записи без ограничений (от регистрации до деклараций, платежей, учета, обеспечения исполнения норм и широкого спектра других вопросов). Значительная часть данных дополнительно представлена в визуальной форме.
- **Комплексный обзор: внешние данные.** Все данные, созданные внешними третьими сторонами, но внесенные в каталог и доступные ведомству. Включают данные как других государственных учреждений, так и партнеров. К внешним данным также относятся данные из ключевых наборов данных, используемых для анализа рисков.
- **Комплексный обзор: дисперсия целевых показателей.** Предварительно рассчитанные отклонения от сопоставимых групп и эталонных наборов данных. Наглядное представление дисперсии за соответствующие налоговые периоды, в основном в разбивке по секторам, отраслям и видам деятельности. Часто включает дисперсию показателей торговли, а также отклонения от экономических показателей (например, темпов роста, инфляции).
- **Комплексный обзор: показатели налогоплательщиков.** Предварительно рассчитанные статистические данные и показатели, включая основные налоговые и финансовые коэффициенты. Логика авторитетности данных может быть консолидирована в одном месте, часто с использованием методов размерного моделирования с предварительным расчетом значений, для обеспечения достоверности результатов и повышения эффективности базовых технологических ресурсов.
- **Комплексный обзор: анализ рисков.** Результаты оценки рисков налогоплательщиков, проводимой аналитиками вручную. Представляет данные, которые носят преимущественно качественный характер, отражая суждения и опыт, накопленный несколькими специалистами за определенное время. В зависимости от имеющихся возможностей также предлагаются последующие шаги для дальнейшего анализа и задачи для выполнения.
- **Комплексный обзор: профили рисков.** Результаты автоматического применения анализа рисков, а также дополнения и комментарии аналитиков. Приоритетное значение уделяется представлению результатов на основе экспертных правил, прогнозной аналитики и сопоставления данных, а также предложениям по дальнейшим действиям. В зависимости от имеющихся возможностей могут отображаться изменения в автоматической оценке рисков, вносимые специалистами вручную.

Источник: персонал МВФ.



Рисунок 5.1 приложения. Факторы, влияющие на поведение налогоплательщиков: BISEP



Фактор BISEP	Элементы оценки
<b>(B) Компания.</b> Информация о налогоплательщике как таковом и о том, является ли он отдельной организацией или частью аффилированной группы.	Объем хозяйственной деятельности, организационно-правовая форма соответствующих структур, местоположение, отрасль и структура капитала.
<b>(I) Отрасль.</b> Характер отрасли, в которой налогоплательщик осуществляет деятельность.	Географический охват, размер, участники отрасли и характер конкуренции, роль ассоциаций, структура затрат, установленные для отрасли нормы и правила.
<b>(S) Социологические факторы.</b> Общее восприятие налогоплательщика в основных регионах, где он ведет свою деятельность.	Общественные нормы и ожидания, степень саморегулирования и репутация налогоплательщика и его руководителей разных уровней.
<b>(E) Экономические факторы.</b> Экономическая ситуация, в которой налогоплательщик в настоящее время ведет свою деятельность, и ожидаемые им будущие условия.	Внутренний или международный рост, внешняя торговля, налоги, процентные ставки, колебания валютных курсов, реформы в области налогово-бюджетной и денежно-кредитной политики.
<b>(P) Психологические факторы.</b> Менталитет, склонности и философия ключевых и влиятельных лиц, определяющие процесс принятия решений.	Цели лиц, принимающих решения, подходы и факторы риска, система управления и взаимоотношения с налоговыми и другими органами.

Источник: концепцию BISEP разработала доктор наук Валери Брейтуэйт в начале 1990-х годов.

Примечание. BISEP – компания, отрасль, социологические, экономические, психологические факторы.

**Принятие формы анализа рисков может облегчить рассмотрение вариантов автоматизации и разработки цифровых инструментов.** Несмотря на то, что потребность в автоматизации аналитики в целом хорошо осознается, дополнительная потребность в инструментах управления профилями не столь очевидна. После принятия и использования соответствующей формы качественные аспекты управления рисками и их значимость становятся более понятны. Это может способствовать продвижению в вопросе развития ИТ-систем.

Для развития ИТ, как правило, рассматривается некоторое сочетание следующих вариантов:

- **Приобретение или усовершенствование модуля рисков.** Расширение концепции модулей, которые могут использоваться для других аспектов администрирования (например, регистрации), в рамках существующих ИТ-платформ, которые, хотя и полезны для определенных целей, могут создавать технические ограничения.
- **Приобретение отдельной системы управления рисками.** Разработка новой ИТ-системы с нуля для удовлетворения широкого спектра потребностей в функциональности, связанной с рисками, включая потребности специалистов по анализу рисков в автоматизации аналитических процессов и инструментальной поддержки.
- **Интеграция коммерческих продуктов.** Использование специализированных технологических платформ и инструментов для работы с инфраструктурой данных (например, озеро данных, хранилище данных), программами статистического анализа, профилирования, риск-ориентированного рабочего процесса и ведения дел.

С целью содействия внедрению анализа рисков и продвижению дискуссий по развитию потенциала в инструментарии к этой справке представлена соответствующая форма. На рис. 5.2 приложения показаны основные элементы формы, в которой для структурирования оценки крупных налогоплательщиков используется модель BISEP.

Рисунок 5.2 приложения. Фрагмент формы профилирования крупных налогоплательщиков из инструментария

	Sources of Information	Key Questions	General Observations	Tax Risk Observations
<p>(B) Business</p> <p><b>The extent and nature of the group's business activities and transactions</b></p> <ul style="list-style-type: none"> <li>• Government registration systems</li> <li>• Business operations</li> <li>• Functional analysis</li> <li>• Value chain analysis</li> <li>• Financial analysis</li> <li>• Annual reports</li> <li>• External research providers</li> <li>• Taxpayer's web site(s)</li> </ul>		<p>What businesses are licensed?</p> <p>What does the taxpayer do to earn its revenue?</p> <p>How does the taxpayer go to market?</p> <p>How does the taxpayer conduct business with the outside world?</p> <p>What assets does the taxpayer use?</p> <p>What risks does the taxpayer take or bear in its business activities?</p>		
<p><b>Entity and group structure</b></p> <ul style="list-style-type: none"> <li>• Annual reports</li> <li>• External research providers</li> <li>• Taxpayer presentations to the tax office</li> </ul>		<p>Do we understand the taxpayer's current legal structure?</p> <p>Is the taxpayer's structure aligned (e.g., legal, organizational, operational)?</p> <p>What types of entities are used (e.g., trusts, partnerships, joint ventures)?</p> <p>What does the taxpayer report by group segment?</p> <p>Is the group structure complex?</p> <p>If the group structure is complex, why is it so?</p> <p>Is there a historical or regulatory reason for the group structure?</p> <p>Does the structure make sense when compared to its business model?</p> <p>Does the structure make sense when compared to its actual operations?</p> <p>What inconsistencies exist between the structure, business model and operations?</p> <p>Have there been any recent changes to the structure? Why?</p> <p>What functional areas does the taxpayer have?</p> <p>What cost centers does the taxpayer have?</p> <p>Is an adequate return being generated from functional areas?</p> <p>Which functional areas are not profitable? Why?</p>		
<p><b>Location and size of the group, including related party dealings</b></p> <ul style="list-style-type: none"> <li>• Annual reports</li> <li>• Income tax return</li> <li>• International dealings schedule</li> <li>• Other tax information</li> </ul>		<p>Where does the taxpayer group operate?</p> <p>What percentage of income is derived nationally?</p> <p>What percentages of tax, profit and sales is made in each jurisdiction?</p> <p>Are any low tax jurisdictions used?</p>		

Элементы оценки, упорядоченные по факторам BISEP (например, "(B)")

Темы для рассмотрения специалистом в форме вопросов

Рабочее пространство для отражения выводов с разных точек зрения

Рекомендации источников информации, полезных для профилирования и анализа

Анализ, помогающий оценить вероятность и последствия наступления рисков

Источник: персонал МВФ.

Примечание. BISEP – компания, отрасль, социологические, экономические, психологические факторы.

## Приложение 6. Отдельная тема. Отбор кандидатов для проведения проверок

Отбор кандидатов для проведения проверок уже долгое время является основой практики управления рисками в налоговом администрировании. В наиболее развитых налоговых администрациях сложные аналитические процессы способствуют оптимизации этой работы в рамках надежных систем управления комплаенс-рисками (УКР). Отражающие кульминацию проб, ошибок и инноваций методы отбора налогоплательщиков, которые все чаще используются для поддержки деятельности в области УКР, имели скромное начало. В рамках аналитики первые попытки часто сводились к использованию данных третьих сторон и разработке надежных методов выявления случаев занижения данных в самостоятельно заполненных декларациях. Сегодня использование аудиторских проверок строится на широком наборе идей, многие из которых становятся более эффективны благодаря росту объема данных и имеющимся методам прогнозирования, позволяющим получать лучшее представление о ситуации в области соблюдения налоговых норм. К этому набору идей относятся принципы и сложившиеся убеждения, состоящие в следующем:

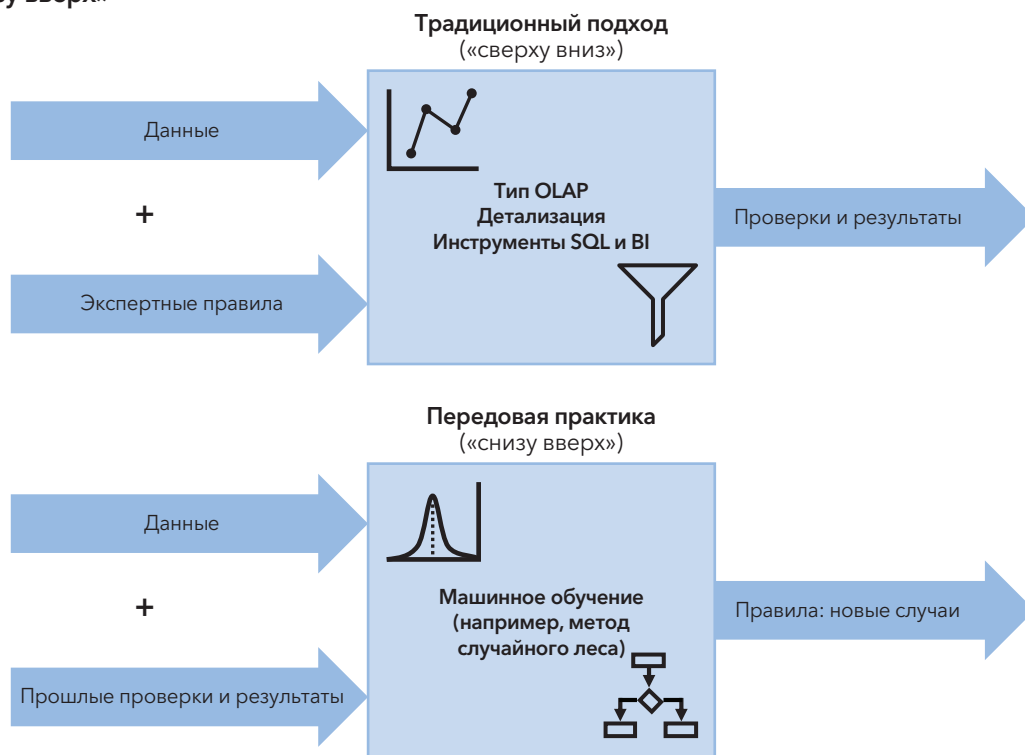
- **Аудит – это прежде всего инструмент, способствующий добровольному соблюдению налоговых норм.** Сегодня аудит, как правило, не рассматривается как инструмент непосредственного получения дохода, а служит ряду целей, включая сдерживание, отслеживание соблюдения норм, выявление случаев несоблюдения норм, информирование и сбор данных.
- **Логика отбора кандидатов для проведения проверок должна быть обоснованной.** Формальная процедура отбора налогоплательщиков должна оформляться документально, основываться на рациональных принципах и последовательно применяться с использованием шагов, которые при необходимости можно обосновать, в том числе при подаче налогоплательщиком апелляции.
- **Централизованный отбор приводит к лучшим результатам.** Современные методики, значительно отличающиеся от первоначальных подходов, при которых аудиторы отбирали кандидатов самостоятельно, по существу, централизуют весь процесс отбора (или его большую часть) с помощью систем и процедур, которые позволяют получать улучшенные результаты.
- **Эффективная централизация учитывает местный опыт.** Во избежание потенциальной чрезмерной зависимости от автоматизации и аналитики при разработке правил оценки рисков для отбора кандидатов для проверок в используемых процессах должен учитываться опыт сотрудников, а по завершении работы должна обеспечиваться обратная связь.
- **Применяемые аналитические процессы должны отражать масштаб деятельности и уровень развития.** В небольших юрисдикциях или в юрисдикциях с низким качеством данных использование базовых методов, таких как простое сопоставление с наборами данных третьих сторон, может быть более практичным, чем использование прогнозной аналитики.
- **Расширенная аналитика наиболее полезна в отношении мелких и средних налогоплательщиков.** Поскольку крупных налогоплательщиков относительно мало и их случаи сложны, имеющиеся данные о прошлых проверках могут представлять ограниченную ценность в качестве исходных данных для алгоритмов прогнозирования, используемых для классификации и отбора кандидатов для проверок.
- **Перед внедрением важно проводить анализ эффективности.** Разрабатываемую логику отбора следует тестировать с помощью результатов прошлых аудиторских проверок для оценки ее точности и получения некоторой степени уверенности в ее эффективности до начала фактического использования.
- **Также очень важен мониторинг логики отбора после ее внедрения.** Прогнозирующие возможности метода отбора могут отклоняться от результатов анализа, проведенного до внедрения модели, в связи

с факторами, не отраженными в результатах прошлых проверок (например, в связи с изменениями экономических условий или непосредственно налоговой политики).

- **Экспертные правила и прогнозная аналитика могут применяться вместе.** Существует множество практических сценариев, при которых простые правила оценки рисков, основанные на опыте, могут быть эффективны, и, пока не будет доказана эффективность применения более сложных методов, их следует регулярно и активно использовать.
- **При использовании аналитики идеальной целью является определение следующего наиболее подходящего кандидата для проверки.** Единый набор интегрированных алгоритмов должен создавать ранжированные списки налогоплательщиков, чтобы в любой момент времени специалисту было ясно, какого налогоплательщика следует проверить следующим (вместо ситуации, в которой аналитики выбирают из списка потенциальных кандидатов с указанием на высокий, средний или низкий уровень риска, внося при этом смещения).

Многие современные взгляды на процесс отбора кандидатов для проверок строятся на различиях между традиционными и передовыми подходами. В обоих случаях аналитика часто опирается на использование правил. Как показано на рис. 6.1 приложения, в традиционном подходе данные сочетаются с правилами, выведенными экспертами. Используя SQL и, все чаще, специализированные модели данных (например, "OLAP"), аналитики применяют эти правила, присваивают каждому из них оценку по балльной шкале и анализируют результаты. По существу, этот подход представляет собой подход «сверху вниз», начинающийся с определения правил. При этом при использовании расширенной аналитики алгоритмы способны самостоятельно выводить правила «снизу вверх».


**Рисунок 6.1 приложения. Отбор налогоплательщиков для проверок по методу «сверху вниз» и «снизу вверх»**



Источник: персонал МВФ.

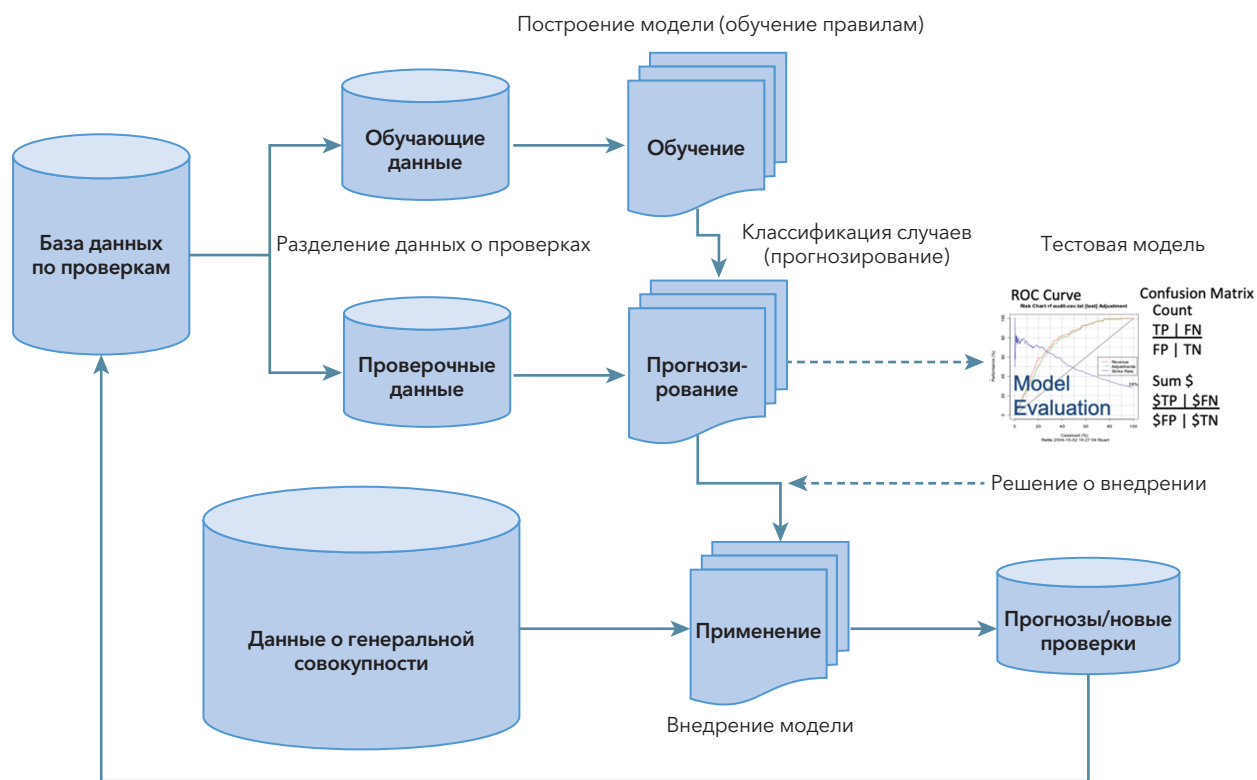
Примечание. BI – бизнес-аналитика; OLAP – Оперативная аналитическая обработка данных; SQL – язык структурированных запросов.

Таблица 6.1. приложения. Традиционный подход и передовая практика

Традиционный подход (отбор кандидатов на основе экспертных правил)	Передовая практика (отбор кандидатов на основе данных)
<p>Профильные эксперты опираются на свой опыт для разработки правил отбора кандидатов, позволяющих выделять клиентов, не соблюдающих нормы, из более широкой генеральной совокупности исходя из конкретных рисков.</p>	<p>Прошлые проверки несоблюдения норм подразделяются на «результативные» (когда было обнаружено соответствующее несоблюдение норм) и «нерезультативные» (когда оно не было обнаружено). После этого набор данных подразделяется на «набор обучающих данных» (используемых для построения аналитической модели отбора кандидатов для проверки) и «набор данных для проверки или подтверждения», позволяющий проверить, работает ли модель.</p>
<p>Эти эксперты обращают внимание на характеристики клиентов, которые, по их мнению, помогают определить, соблюдает ли клиент налоговые нормы.</p>	<p>При помощи профильных экспертов специалист по анализу данных определяет характеристики клиентов, которые, как представляется, связаны с несоблюдением налоговых норм. Эти характеристики тестируются с использованием имеющейся статистики с целью выяснить, действительно ли они связаны с несоблюдением норм.</p>
<p>Созданные правила часто уточняются с течением времени, чтобы повысить коэффициент результативности проверок. Для оценки результатов часто используются сводные статистические данные, такие как коэффициент результативности проверок и величина средней корректировки. Это создает значительные трудности в случае асимметричных распределений.</p> <p>Разработанные правила, как правило, взвешиваются на субъективной основе для оценки риска клиентов по балльной шкале с целью определения приоритетности кандидатов для проверки.</p>	<p>Далее выбирается набор обучающих данных и, по существу, строится регрессия целевого набора результативных проверок для создания алгоритма оценки рисков, позволяющего оптимизировать вероятность верного прогноза на основе этих данных.</p> <p>После этого выбирается алгоритм машинного обучения, эффективность которого тестируется на отдельном наборе проверочных данных с помощью таких методов, как ROC-кривая и матрица несоответствий.</p>
<p>Гипотетический пример:</p> <ul style="list-style-type: none"> <li>• Если клиент заявляет вычет по расходам, связанным с работой, в размере &gt; 1 500 долл. и расходам на униформу в размере &gt; 500 долл., то балл риска = 4.</li> <li>• Если клиент заявляет вычет по расходам на транспортное средство в размере &gt; 5 000 долл., то балл риска = 2.</li> <li>• Если клиент заявляет вычет по расходам на самообразование в размере &gt; 2 000 долл., то балл риска = 3.</li> <li>• Баллы риска клиента суммируются для получения общей оценки риска.</li> <li>• Клиенты с общим баллом риска &gt; 8 отбираются для проверки.</li> </ul>	<p>Полученные правила взвешиваются алгоритмом для получения оценки риска клиента с целью определения приоритетности кандидатов для проверки:</p> 
<p>Преимущество правил, разработанных таким образом (сверху вниз), заключается в том, что они известны профильному эксперту и легче поддаются интерпретации, но в целом они не являются оптимальными.</p>	<p>Разрабатываемые таким образом (снизу вверх) правила иногда вызывают удивление у профильных экспертов, а их понимание и интерпретация могут требовать определенных усилий (однако для обеспечения объяснительной способности результатов сложной модели, имеющей более высокий коэффициент результативности проверок, как правило, можно модифицировать простое дерево решений).</p>
<p>При сложении баллов риска получается централизованное кластерное распределение рисков. Умножение баллов риска приводит к логнормальному распределению. Влияние на порядок ранжирования минимально.</p>	<p>Правила, разрабатываемые с помощью таких подходов на основе данных, как правило, более эффективны по сравнению с правилами, выводимыми на основе субъективного мнения специалистов об отличительных характеристиках и весах, которые им следует присвоить.</p>
<p>Источник: персонал МВФ. Примечание. ROC-кривая – кривая соотношения истинных и ложных характеристик.</p>	

В настоящее время передовая практика отбора кандидатов для проверок опирается на прогнозирующую аналитику и алгоритмы машинного обучения, которые часто основываются на единой базовой концепции. Как показано на рис. 6.2 приложения, результаты прошлых проверок служат основой для алгоритмов прогнозирования будущего поведения. При этом в применяемые алгоритмы закладывается не выраженное в явной форме допущение о том, что результаты прошлых проверок в достаточной мере отражают поведение налогоплательщиков в разные периоды времени. По множеству причин (например, в связи со значительными изменениями в налоговой политике) это допущение не всегда является верным. Тем не менее, описанная концепция позволяет получить результаты, которые можно протестировать перед внедрением и сравнить с другими подходами (например, использованием экспертных правил), чтобы объективно оценить, какой из них обладает наилучшими прогностическими возможностями с учетом имеющихся данных.

Рисунок 6.2 приложения. Отбор налогоплательщиков для проверок с использованием прогнозной аналитики



Источник: персонал МВФ.

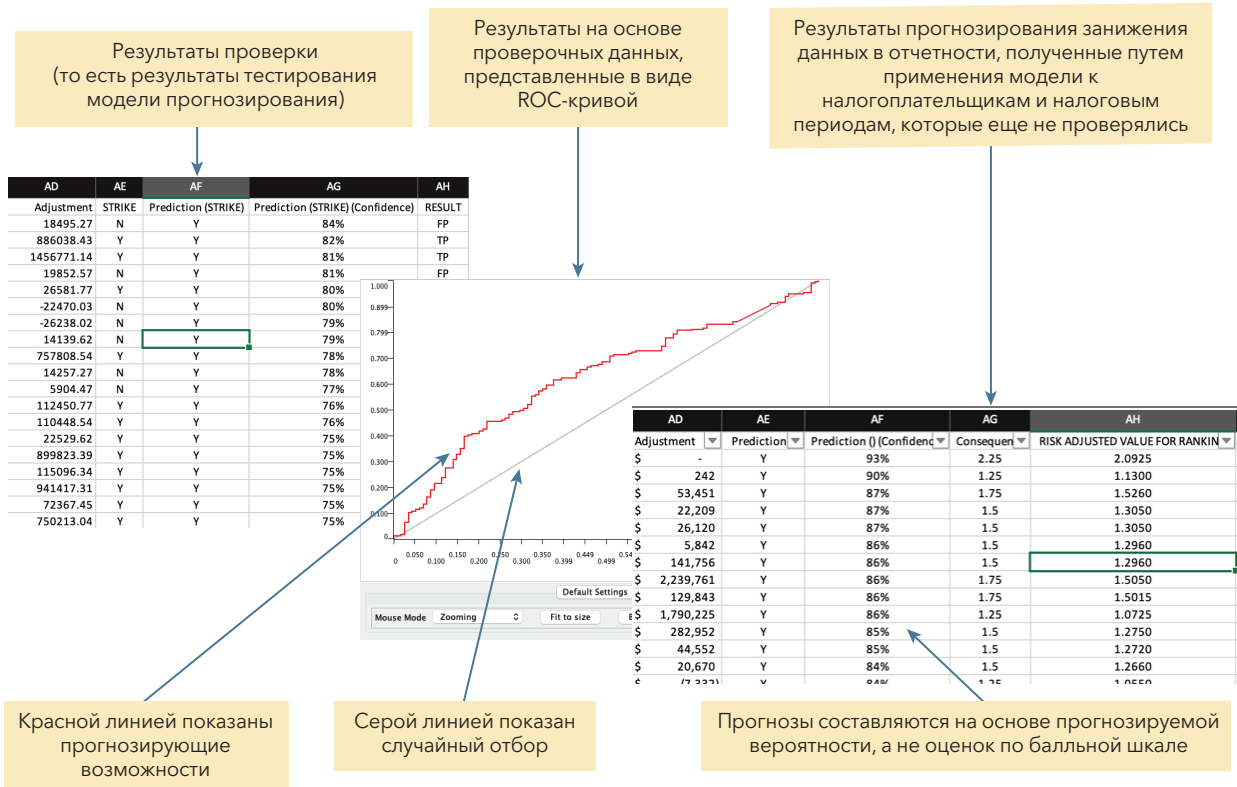
**Несмотря на существующие различия, основные методы прогнозной аналитики, как правило, опираются на одни и те же или схожие компоненты, в том числе следующие:**

- **База данных по проверкам.** Содержит результаты проверок и точки данных, на которые опираются алгоритмы для составления прогнозов (часто это доначисленные по результатам аудиторской проверки суммы).
- **Обучающие данные.** Относительно крупная случайная выборка данных из базы данных по проверкам (от 60 до 80 процентов), используемая алгоритмом для выявления закономерностей и правил.
- **Обучение.** Алгоритм, обнаруживающий закономерности и правила в данных о прошлых проверках на основе заданной специалистом конфигурации (например, обучающий алгоритм случайного леса).
- **Проверочные данные.** Менее крупная случайная выборка данных из базы данных по проверкам (от 20 до 40 процентов), используемая для проверки точности прогнозов, выполненных алгоритмом.
- **Прогнозирование.** Алгоритм, использующий закономерности и правила, выявленные в ходе обучения, для составления новых прогнозов с использованием данных о применении и данных о генеральной совокупности (то есть налоговых данных о налогоплательщиках, которые еще не проверялись).
- **Тестовая модель.** Рабочий процесс, использующий проверочные данные и методы измерения эффективности (ROC-кривые) для определения прогнозирующих возможностей.
- **Применение.** Модели, часто используемые только в аналитических инструментах, могут применяться непосредственно на ИТ-платформах, поддерживающих основную операционную деятельность.
- **Данные о генеральной совокупности.** Налоговая база данных, включающая информацию о налогоплательщиках, которые еще не выбирались для проверки по оцениваемым вопросам и периодам.
- **Прогнозы/новые проверки.** После составления прогнозов соответствующие данные вносятся в базу данных по проверкам, на их основе определяются задачи для выполнения, и вся информация отслеживается, с тем чтобы использовать результаты при проведении нового анализа.

**Все чаще разрабатываются компьютерные программы расширенной аналитики для специалистов, не имеющих навыков программирования или владеющих лишь базовыми навыками, что значительно повышает доступность этих методов.** Если раньше для многих ведомств отсутствие навыков программирования было препятствием, то сегодня специалисты по анализу данных могут использовать все описанные передовые концепции, не имея навыков программирования. Хотя поддержка со стороны персонала, владеющего навыками работы с SQL, как правило, по-прежнему необходима (обычно требуется помощь ИТ-персонала по извлечению необходимых наборов данных), после получения требуемых данных их можно анализировать с использованием полного набора методов и алгоритмов посредством визуального проектирования и настройки соответствующих моделей и рабочих процессов.

**На рис. 6.3 приложения представлены результаты применения формы отбора налогоплательщиков, представленной в инструментарии, сопровождающем настоящую справку.** Хотя общая осведомленность и понимание потенциала расширенной аналитики могут быть полезны, разработка и использование формы не требуют владения языками программирования.

**Рисунок 6.3 приложения. Результаты применения формы налогоплательщиков для проверок из инструментария**



Источник: персонал МВФ.

Примечание. ROC-кривая – кривая соотношения истинных и ложных характеристик.



## Литература

- Aslett, Joshua, et al. 2024. IMFx: VITARA - Information Technology and Data Management. edX. <https://www.edx.org/school/imfx>
- Betts, Sue. 2022. "Revenue Administration: Compliance Risk Management Framework to Drive Revenue Performance." IMF Technical Note 2022/005, International Monetary Fund, Washington, DC.
- Brondolo, John, Annette Chooi, Trevor Schloss, and Anthony Siouclis. 2022. "Compliance Risk Management: Developing Compliance Improvement Plans." IMF Technical Note 2022/001, International Monetary Fund, Washington, DC.
- Chooi, Annette, and Jonathan Leigh Pemberton. 2023. IMFx: VITARA - Compliance Risk Management. edX. <https://www.edx.org/learn/economics/the-international-monetary-fund-vitara-compliance-risk-management>
- Cotton, Margaret, and Gregory Dark. 2017a. "Use of Technology in Tax Administrations 1: Developing an Information Technology Strategic Plan (ITSP)." IMF Technical Note 2017/001, International Monetary Fund, Washington, DC.
- Cotton, Margaret, and Gregory Dark. 2017b. "Use of Technology in Tax Administrations 2: Core Information Technology Systems in Tax Administrations." IMF Technical Note 2017/002, International Monetary Fund, Washington, DC.
- Cotton, Margaret, and Gregory Dark. 2017c. "Use of Technology in Tax Administrations 3: Implementing a Commercial-Off-The-Shelf (COTS) Tax System." IMF Technical Note 2017/003, International Monetary Fund, Washington, DC.
- Organisation for Economic Co-operation and Development (OECD). 2016. *Advanced Analytics for Better Tax Administration: Putting Data to Work*. Paris: OECD Publishing.





## PUBLICATIONS

Налоговое администрирование: основные возможности  
аналитики для управления комплаенс-рисками (Russian)

TNM/2024/01



9 798400 272240