

22 октября 2024

Истинный data-driven, или как избежать «старения» данных?

Сегодня только около 20% крупных российских компаний применяют полноценный data-driven подход. В основном это ведущие ритейлеры и финтех, поскольку именно для этих рынков характерна высокая конкуренция, требующая непрерывного анализа ситуации и оперативных управленческих решений, основанных на актуальных данных. Остальные компании так или иначе работают с Big Data, но назвать это системным data-driven подходом нельзя. Разбираемся, что включает в себя это понятие и какие ошибки чаще всего допускает бизнес при работе с данными.

До появления data-driven подхода были распространены highest paid person's opinion (HiPPO) подход (принятие решений на основе мнения и опыта руководства) и best practices подход (принятие решений на основе общепринятых, лучших отраслевых стандартов). В современной экономике, где внешние и внутренние условия постоянно меняются, подход HiPPO зачастую малоэффективен: мнение одного человека, в данном случае руководителя, субъективно, часто основывается на интуиции, а ошибка может стоить компании очень дорого. Что касается best practices подхода, который основан на стандартизации всех процессов, то фактически это усредненные



данные в определенной области. Однако он тоже годится далеко не для каждой ситуации – условно, то, что работало вчера, может не сработать сегодня; то, что применимо в одной компании может не подойти другой, даже если речь про одну и ту же отрасль.

Общая цифровизация, а также внедрение веб-сервисов и мобильных приложений, собирающих внешние данные, послужили сильным толчком к развитию data-driven подхода, главный признак которого – регулярный автоматизированный сбор и анализ данных. Полноценный data-driven подход включает идею, проверку гипотез, интерпретацию экспериментов, принятие решений, изучение последствий. Например, компания хочет сократить время, который затрачивает пользователь на поиск товара, и решает внести изменения в интерфейс приложения. Чтобы понять, будет ли успешен такой эксперимент, гипотеза сначала проверяется на ограниченном числе пользователей. Если эксперимент проходит удачно, тогда изменения становятся доступны всем пользователям. Затем компания мониторит метрики внесенных изменений, которые сравниваются с метриками старого функционала. Таким образом, происходит постоянный мониторинг внешних и внутренних изменений, а данные не успевают устареть. На практике многие компании работают с данными, однако делают это не системно, допуская распространенные ошибки, которые мешают в полной мере извлекать пользу из Big Data.

3 САМЫЕ ЧАСТЫЕ ОШИБКИ ПРИ РАБОТЕ С ДАННЫМИ

ОШИБКА 1. НИЗКИЙ УРОВЕНЬ АВТОМАТИЗАЦИИ = «ГРЯЗНЫЕ» ДАННЫЕ

Для того, чтобы получать реальную ценность от data-driven подхода, важно иметь «чистые» данные, на которых в дальнейшем можно строить качественный достоверный анализ. Любая погрешность в исходных данных, влечет за собой дальнейшие ошибки при их интерпретации. Здесь существует простая закономерность: чем меньше процессов в компании автоматизировано, тем больше вероятность получить некорректные данные. Например, если информация заносится в систему вручную, риск ошибки намного больше из-за человеческого фактора. Кроме того, если в одном случае данные собираются автоматически, а другом вручную, то исчезает возможность сравнивать актуальные показатели. Наконец полноценному использованию data-driven подхода часто мешает разрозненность накопленных данных. Решением может быть комплексная автоматизация бизнеса: в этом случае компания будет получать максимально полную картину.

ОШИБКА 2. ОТСУТСТВИЕ ИНТЕГРАЦИИ = РУЧНАЯ АГРЕГАЦИЯ

Как правило в рамках одной компании существует несколько различных информационных систем, работающих с разными подразделениями, производственными и бизнес-процессами. Кроме того, бизнес часто



использует инструменты для сбора внешних данных. Зачастую все эти источники функционируют независимо друг от друга, что затрудняет автоматический сбор и агрегацию данных. Проблема решается путем интеграций всех систем в единое хранилище данных – создаются автоматизированные аналитические механизмы для очистки некачественных входных данных в составе пайплайнов обработки, что минимизирует вероятность ошибки при дальнейшем анализе. Также агрегации и расчеты входят в пайплайны обработки данных для получения сводной аналитики. Далее готовые результаты отправляются в конечную точку вывода.

ОШИБКА 3. ОТСУТСТВИЕ ВИЗУАЛИЗАЦИИ СВОДНОЙ ОТЧЕТНОСТИ = ДОЛГИЙ ПОИСК ДАННЫХ

Мало обрабатывать данные, нужно еще и визуализировать их таким образом, чтобы они были понятны и быстро доступны для конечного пользователя. Время, которое компания тратит на сбор регулярной отчетности вручную – это время, на которое она откладывает принятие своих управленческих решений. Кроме того, отсутствие корректной и эффективной визуализации мешает «провалиться» на более глубокий уровень детализации. Многие решают эту проблему огромным числом табличных файлов, но тогда для поиска нужных данных иногда требуется еще больше времени, чем на сбор самой отчетности. Избежать этой ошибки помогают интерактивные дашборды, настроенные исходя из целей и контекста.

КАК ВЫГЛЯДИТ ИДЕАЛЬНЫЙ DATA-DRIVEN ПОДХОД?



Понять, применяет ли бизнес полноценный data-driven подход, достаточно просто: необходимо посчитать, сколько в среднем сотрудник тратит на сбор и обработку данных, а затем проанализировать, насколько эти данные успели устареть. Все вышеперечисленные ошибки приводят к тому, что пока сотрудники заняты исправлением ошибок данные успевают «испортиться». Неэффективная работа с данными, становятся не драйвером, а проблемой, связанной с постоянным поиском и исправлением ошибок, ручной обработкой и работой с уже устаревшими данными. В таких условиях некорректно говорить, что компания применяет в работе data-driven подход, скорее она находится только на пути к его применению.

Если же процесс сбора данных автоматизирован, настроены автоматизированные аналитические механизмы обработки данных, присутствует качественная визуализация сводной отчетности и эти процессы требуют минимального участия человека (поддержка системы, улучшение алгоритмов обработки данных), то компания может извлекать из Big Data максимальную пользу. В этом и основная ценность data-driven подхода – достоверная и актуальная информация беспрепятственно поступает от источника к конечному потребителю за минимальное количество времени.