9лента

Reliable ML: прикладной подход к управлению внедрением инициатив продвинутой аналитики

Ирина Голощапова Head of Data Science

Обо мне

- Head of Data Science в ЛЕНТА
- ► Консультант в Skolkovo School of Management
- Организатор в Open Data Science Community
- ▶ PhD in Economics
- **▶** B DS c 2009 r.:
 - Статистический анализ и моделирование в ЦМАКП
 - DS в рисках, макроэкономике и оптимизации бизнес-процессов в X5 Group
 - DS Lead в Manchester University в социальнозначимых проектах
- Telegram-channel Reliable ML:

https://t.me/reliable_ml



@irina_goloshchapova



i.o.goloshchapova@gmail.com





















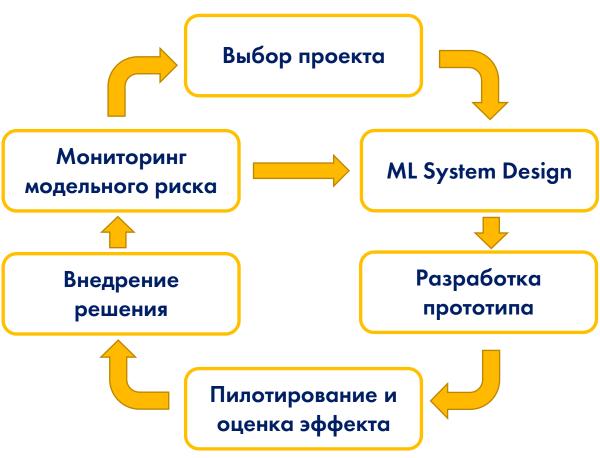
Что такое Reliable ML?

Фреймворк по внедрению и развитию продвинутой аналитики

Как сделать результат работы data science команд применимым в бизнес-процессах компании и приносящим компании прибыль:

- Выбор направлений для внедрения продвинутой аналитики
- Проектирование и реализация решения продвинутой аналитики в контексте бизнес-процесса его применения
- Учет причинно-следственных связей и необходимости интерпретации моделей для пользователя в ML алгоритмах
- Статистически достоверная оценка пилотных экспериментов
- Управление внедрением решения и модельным риском



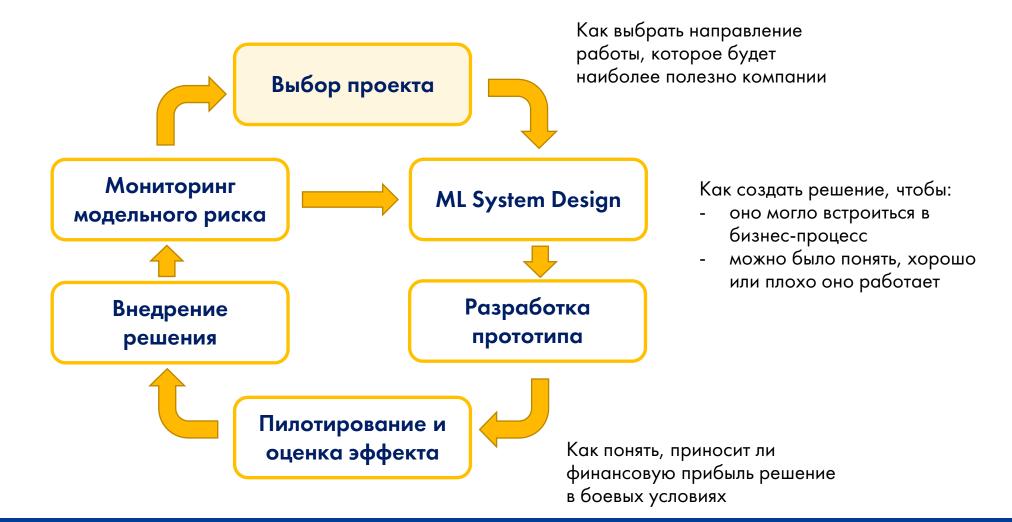




Фреймворк Reliable ML

Где внедрение МL даст наибольшую отдачу?

Как сделать из прототипа стабильное боевое решение и определить, когда его нужно дорабатывать



Reliable ML: Какие инициативы продвинутой аналитики реализовывать?

Ключевые принципы выбора инвестиционных инициатив в области продвинутой аналитики

ACTIONABLE

/

Сложность реализации инициативы средствами продвинутой аналитики



Применимость инициативы для текущих бизнес-процессов

MEASURABLE



Для инициативы возможно проведение пилотного эксперимента и корректная оценка ее эффекта на ключевые бизнес-показатели компании

IMPACT



Для инициативы возможно рассчитать ожидаемый эффект на ключевые бизнеспоказатели компании



Эффект является материальным с точки зрения PnL компании



Эффект является достижимым в ближайшие 12 месяцев (Quick-Wins First)

BUSINESS PRIORITY



Оценка приоритетности выполнения инициативы со стороны бизнесподразделений

POSITIVE BUSINESS CASE



Ожидаемый эффект от реализации инициативы превышает затраты на проект



Инициатива может быть встроена в бизнес-процессы в ближайшие 12 месяцев (Quick-Wins First)



Продвинутая аналитика в Ленте

Алгоритмы & Сервисы & Аналитика & Пилоты & Валидация

Оптимизация ассортимента

Какие товары завезти в магазины?

Прогноз спроса

Какой будет спрос при различных предпосылках?

Прогноз выручки новых ТК

Сколько будет зарабатывать новый магазин?



Управление промо

Какой установить промо-календарь?

Динамическое ценообразование

Какие цены установить?

Единая методика АБ-тестирования

Как оценить эффект от пилота?

Лента.Онлайн

Улучшение пользовательского опыта в поиске и рекомендациях

Монетизация данных

Обезличенная аналитика и модели по запросу внешних организаций

Business Intelligence

Поддержка принятия решений: ВІ-отчёты, автоматизация повторяющихся расчетов

Клиентская аналитика

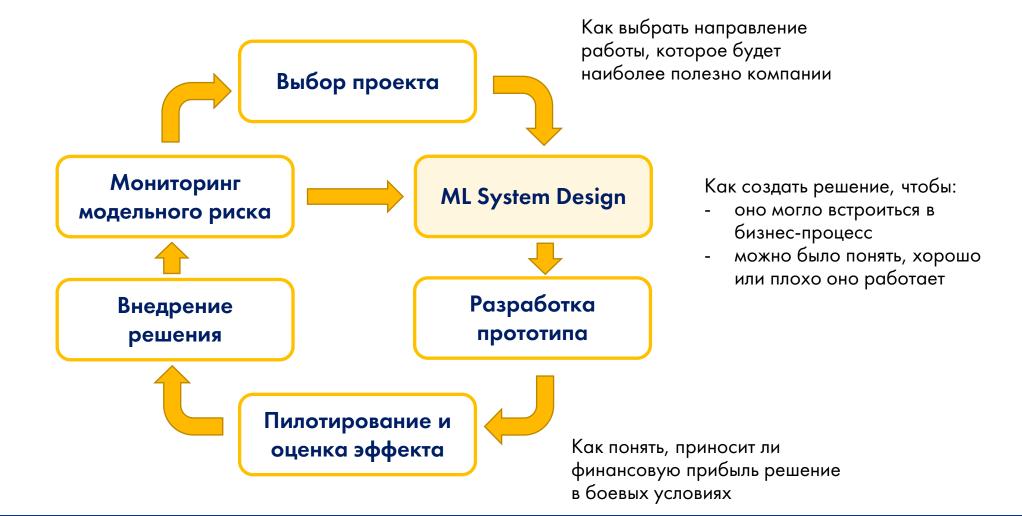
Поддержка действий КМ и отдельные AdHoc задачи



Фреймворк Reliable ML

Каким должен быть дизайн системы машинного обучения?

Как сделать из прототипа стабильное боевое решение и определить, когда его нужно дорабатывать



Как ML-модель встроится в текущий бизнес-процесс инициативы, чтобы приносить пользу?

ML System Design – бизнес-аспект

- От простого к сложному
 - ✓ Сначала простой и эффективный baseline, затем последовательное усложнение с ростом коммерческого эффекта
- Бизнес-процессы компании, связанные с использованием ML-модели, оказывают критическое влияние на дизайн ML системы
 - ✓ Какое место инициатива занимает в общей карте бизнес-процессов компании?
 - ✓ Как ML-модель встроится в текущий бизнес-процесс, чтобы приносить пользу?
- Применимость модели для бизнес-процесса для получения коммерческого эффекта главное на каждом этапе
 - ✓ Baseline этап помогает конкретизировать БТ целевого продукта и определить точки особого внимания для его разработки (данные, алгоритмы, результат)
 - ✓ Метрики качества модели при ее построении должны быть привязаны к процессу применения модели

От простого к сложному

Виды продвинутой аналитики

Предписательная Оптимизация Что именно и когда нужно сделать, аналитика чтобы было хорошо? Прогнозная Причинно-следственный анализ Что будет, если я что-то поменяю в аналитика системе? Прогноз Что будет? Описательная **Statistical Inference** Это произошло повсеместно? аналитика Факторный анализ Из чего складывается это изменение? Ретроспективный анализ Что произошло? Сложность

Ценность

От простого к сложному

Управление промо в Ленте: последовательное развитие продукта для получения максимального эффекта

Рост выручки от Предписательная Оптимизация Эффект 3 улучшения подбора Какой установить промоофферов аналитика календарь, чтобы эффект был максимален? Пилот по Прогнозная Эффект 2 What-if анализ стадии 3 Пилот по Каков будет эффект с учетом внешних аналитика стадии 2 факторов в будущем и как мы можем им управлять? Эффект 1.2 **Ретроспективный** Модельный подход Пилот по Более точная оценка эффекта в анализ стадии 1.2 прошлом с учетом влияния внешних факторов Аналитический подход Эффект 1 Пилот по Каков эффект промо в стадии 1.1 прошлом? Стадия проекта Стадия 1 Стадия 2 Стадия 3

Как ML-модель встроится в текущий бизнес-процесс, чтобы приносить пользу?

Изменение ассортимента в магазине

Kak Data Scientists представляют себе управление ассортиментом в магазинах в продуктовом ритейле в начале проекта?

Рекомендации по товарам для каждого магазина



Новый ассортимент и счастливые клиенты в каждом магазине

Как ML-модель встроится в текущий бизнес-процесс, чтобы приносить пользу?

Как в продуктовом ритейле в реальности управляют ассортиментом?



12

магазине

^дКак ML-модель встроится в текущий бизнес-процесс, чтобы приносить пользу?

Даже самый умный искусственный интеллект по оптимизации ассортимента не принесет прибыли, если не учитывать критические блоки бизнес-процесса





Какое место инициатива занимает в общей карте бизнеспроцессов компании?

Управление промо в Ленте: Бизнес-процесс – неотделимая часть управления жизненным циклом товара



Какое место инициатива занимает в общей карте бизнес-процессов компании?

Управление промо в Ленте: Единая методика прогноза спроса – основа системы

Единые компоненты в модели прогноза спроса создают фундамент для слаженной работы системы и выявления глобального оптимума

Прогноз спроса на каждый товар в каждом ТК

Кросс-эффекты спроса на товар

ГАЛО эффект

Каннибализация и товары-субституты

Эффект бренда

Товары KVI

Эластичность спроса по цене

Эластичность спроса по регулярной цене Эластичность спроса по промо-цене Эластичность спроса по рег. цене на другие товары

Эластичность спроса по промо цене на другие товары

Потребитель и внешняя среда

Характеристики клиентских сегментов

Сезонность

Характеристики локаций: трафик, конкуренты, инфраструктура

Внешняя среда: макроэкономика, погодные условия

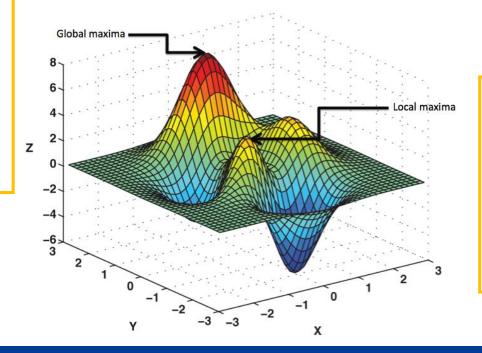
Какое место инициатива занимает в общей карте бизнеспроцессов компании?

Глобальный оптимум возможен только при слаженной работе систем управления жизненным циклом товара

В 2018 г. Gartner опубликовал отчет Market Guide for Unified Price, Promotion and Markdown Optimization Applications, где обозначил, что будущее – за объединенными системами управления промо, товародвижением и ценообразованием

Глобальный максимум

наибольшая выгода для всей системы в целом – определяется по сумме результатов для каждого звена цепи управления жизненным циклом товара

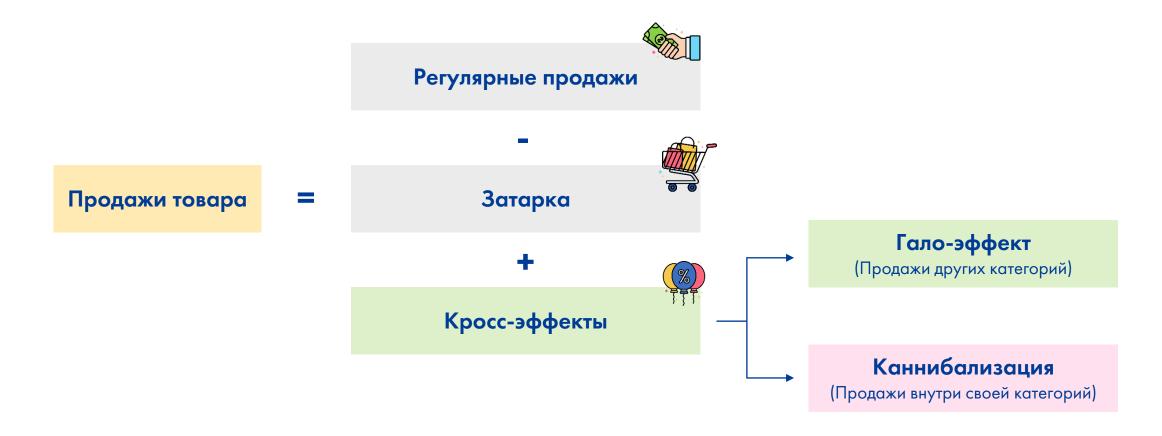




наибольшая выгода для одного продукта (например, промо) – безотносительно результата для ассортимента, ценообразования или товародвижения

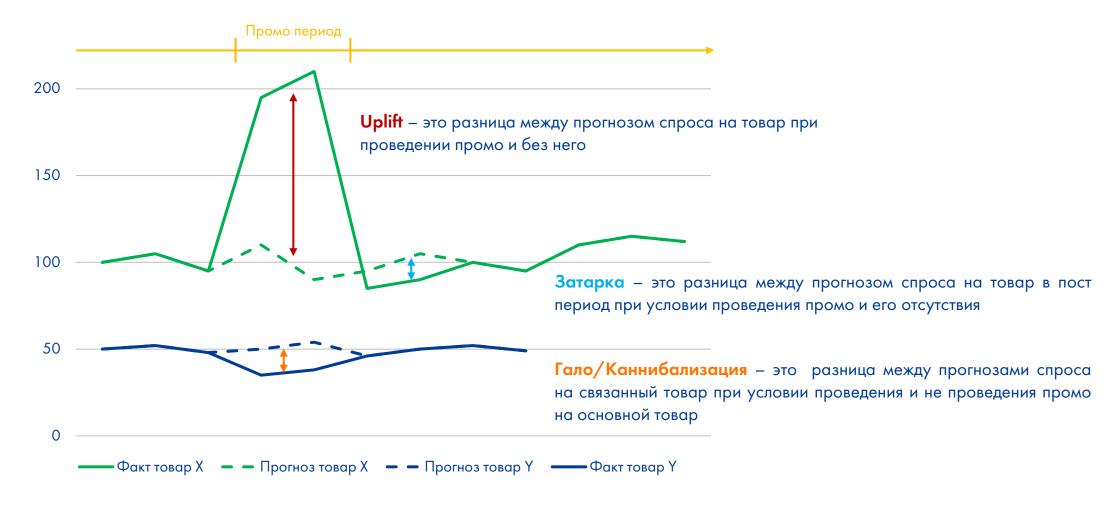
Применимость модели для бизнес-процесса для получения коммерческого эффекта – главное на каждом этапе

Управление промо в Ленте: Ретроспективный анализ промо – аналитический подход



Применимость модели для бизнес-процесса для получения коммерческого эффекта – главное на каждом этапе

Управление промо в Ленте: Ретроспективный анализ промо – модельный подход



Применимость модели для бизнес-процесса для получения коммерческого эффекта – главное на каждом этапе

Управление промо в Ленте: Метрики качества для разработки решения

Аналитический подход

Качественная оценка

- Экспертная оценка релевантности результатов
- Применимость для бизнес-процесса управления промо



Модельный подход

Количественные метрики

Релевантность результатов:

- Качество прогноза спроса по товарам/группам товаров
- Advanced-уровень метрики, касающиеся методов causal inference

Применимость для бизнес-процесса:

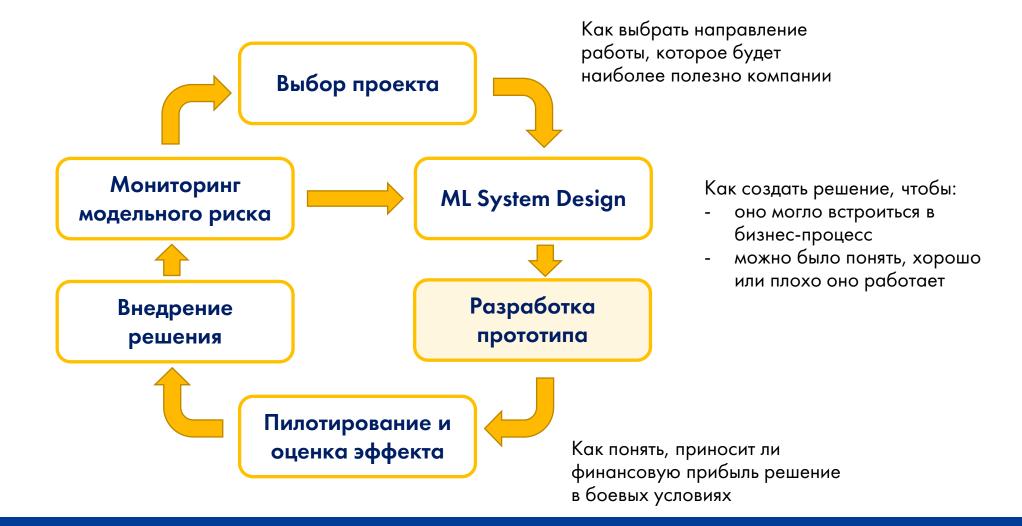
Качество ранжирования результата



Фреймворк Reliable ML

Какие ключевые этапы моделирования должны быть реализованы для получения качественного ML решения?

Как сделать из прототипа стабильное боевое решение и определить, когда его нужно дорабатывать



Какие ключевые этапы моделирования должны быть реализованы для получения качественного ML решения?

Шаблон для разработки ML-решения



Найти вовлеченную и компетентную команду

Какие ключевые этапы моделирования должны быть реализованы для получения качественного ML решения?

Шаблон для разработки ML-решения



2



Имплементация и оценка эффекта



I. Подготовка и препроцессинг данных

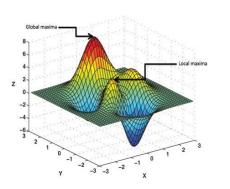
- Расчет факторов для моделирования
- Учет бизнес-ограничений

II. Моделирование

Обучение прогнозных моделей и подготовка факторов для тестового расчета

III. Оптимизация

- Подготовка и реализация симуляций для прогноза
- Оценка симуляций и оптимизация эффекта
- Подготовка финального отчета



IV. Бизнес-проверка

- Бизнес-анализ рекомендаций и внесение экспертных знаний
- Переоценка эффектов с учетом внесенной экспертизы

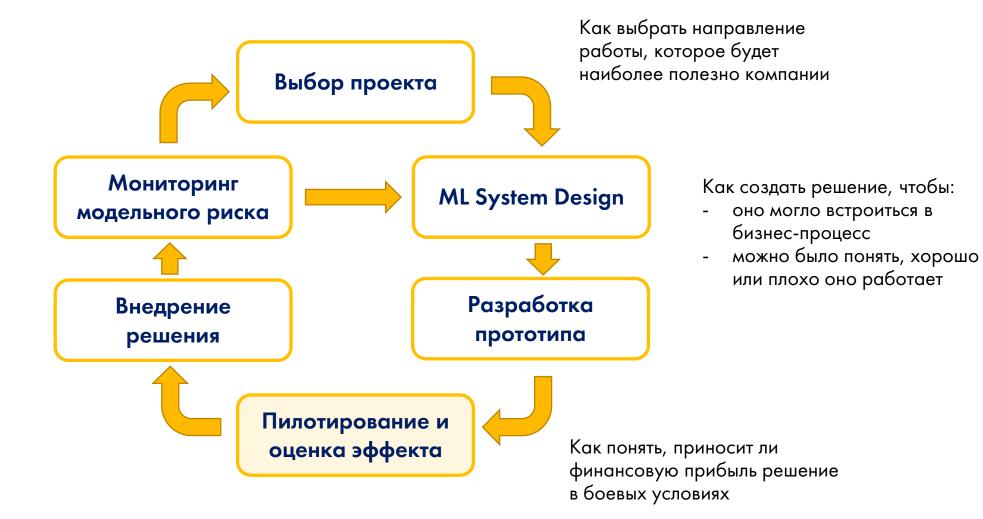




Фреймворк Reliable ML

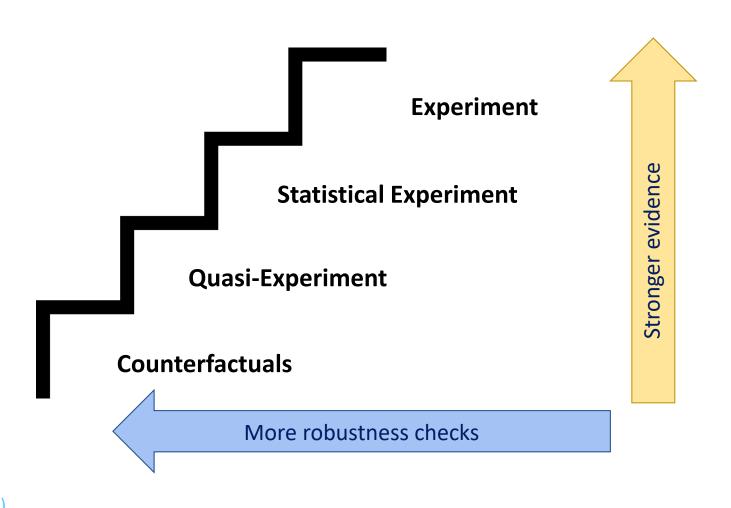
Как – с учетом бизнес-ограничений - мы можем оценить эффективность работы нашей модели?

Как сделать из прототипа стабильное боевое решение и определить, когда его нужно дорабатывать



Как корректно оценить эффективность внедрения инвестиционной инициативы?

Доказательная лестница методов причинно-следственного анализа (Causal Inference Evidence Ladder)



Source: A.Rebecq (2020)



^дКак – с учетом бизнес-ограничений - мы можем оценить эффективность работы нашей модели?

Единая методика оценки эффективности инвестиционных инициатив – ключевой стрим для поддержки цифровой трансформации компаний

Кажется, что оценить эффект просто...



- Выбираем пилотные объекты и похожие на них контрольные
- Сравниваем среднее пилота и контроля после и до пилота
- Получаем реальную оценку эффекта проекта/инициативы

На самом деле, нет



- Простое сравнение средних не работает
- Не можем отличить эффект от
 - случайности
 - влияния внешних факторов
 - специфики выбранных в пилот и контроль объектов
 - манипуляций
- Не знаем, повторится ли найденный эффект при новом эксперименте
- Делаем некорректный вывод о ролл-ауте инициативы и теряем много денег и времени

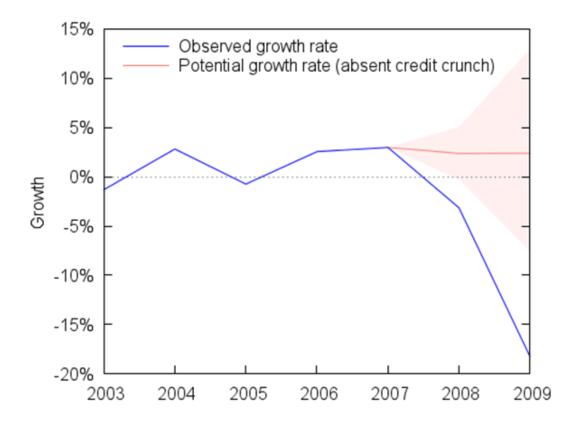


Как – с учетом бизнес-ограничений - мы можем оценить эффективность работы нашей модели?

Что делать, если АБ-тестирование не работает

Решение

Необходимо использовать группу методов причинно-следственного анализа – Counterfactual Analysis





Зачем внедрять единую методику оценки инвестиционных инициатив?

Прямой эффект от внедрения методики составляет от 3% выручки компании

Единая методика оценки эффективности инвестиционных инициатив

Оптимизация инвестиционного потока

- Вовремя понимаем, какие инициативы стоит остановить, а в какие продолжать вкладывать
- Рост дохода от ролл-аута проектов
- Учитываем степень неопределенности при оценке проектов

Оптимизация оплаты работы консультантов и поставщиков IT-решений

- Успех проектов внешних контрагентов тестируется по единой методике
- Оплата переменной части договоров (success fee) производится только в случае объективного успеха проекта

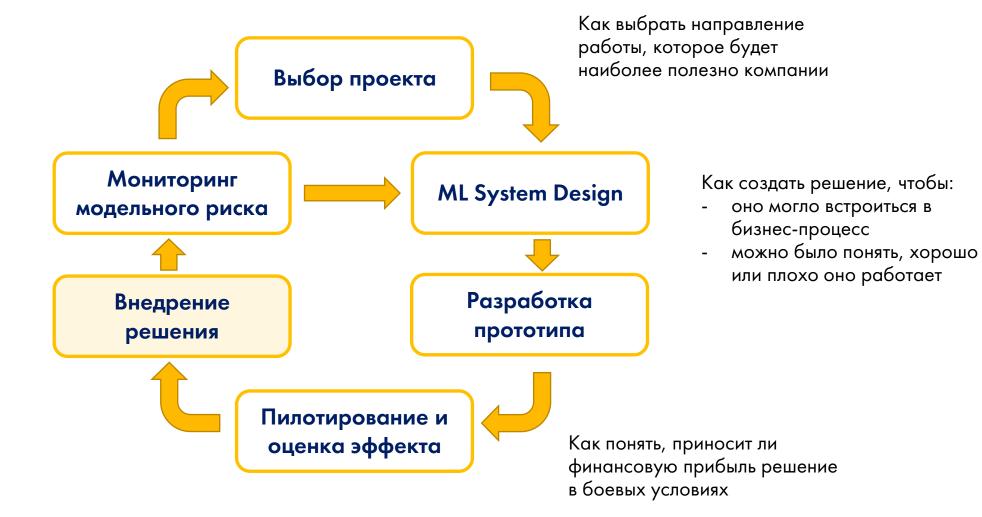
Эффективное управление цифровыми рисками

- Быстрый и прозрачный бизнес-процесс
- Экономия времени, отсутствие двойных стандартов/дублирования в оценке проектов
- Понятная бизнесу методика и SLA

Фреймворк Reliable ML

Как организовать вывод модели в продуктив?

Как сделать из прототипа стабильное боевое решение и определить, когда его нужно дорабатывать



Как организовать вывод модели в продуктив?

Требования к внедрению решения – бизнес-аспекты

■ Доверие к результату

✓ Принятие решений моделью должно быть понятно бизнесу как в целом (global interpretation), так и на отдельных примерах (local interpretation)

■ Применимость модели в реальных условиях

- ✓ Реальные условия всегда отличаются от тех, на которых строилась модель
- ✓ Понятность модели конечному пользователю в продуктиве снижает риск некорректного применения модели на новых данных/при сложных кейсах/в меняющейся среде

Информативность для бизнес-процесса

- ✓ Результат работы модели должен быть информативным для бизнеса
- ✓ Для конечного пользователя должно быть понятно не только как модель пришла к результату, но и как этот результат должен быть использован и почему

Доп. материалы: Интерпретируемость моделей машинного обучения и их представление бизнесу

Как организовать вывод модели в продуктив?

Требования к внедрению решения – технические аспекты

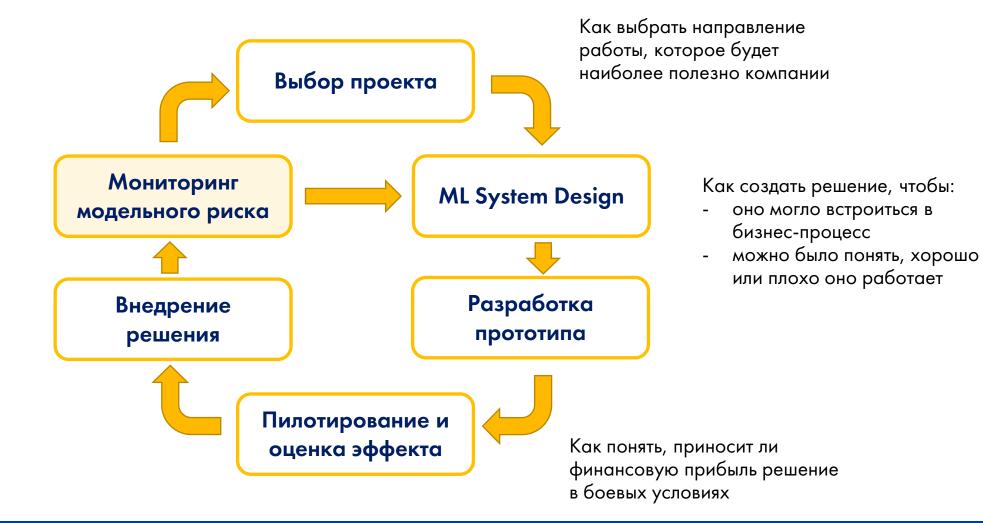
- Отсутствие разрыва DEV-PROD в данных, используемых моделью
 - ✓ Факторы модели должны быть одинаковы при обучении модели и ее работе в продуктиве построение feature store для работы над всеми production-моделями компании
- Архитектура ML-сервисов, отвечающая принципам классической разработки
 - ✓ Сервисы, обеспечивающие работу ML-модели, и их архитектура должны соответствовать ключевым принципам и свойствам классической разработки: модульная структура, стиль, тесты, среда разработки, документация
- Документация
 - ✓ Документация должна включать:
 - ✓ Бизнес-постановку и технические аспекты
 - ✓ Текущее состояние модели, ее предшествующие версии и эксперименты по улучшению
 - ✓ Логику построения и работы как модели, так и production-сервисов

Доп. материалы: Построение Feature Store в компании Лента

Фреймворк Reliable ML

Как организовать вывод модели в продуктив?

Как сделать из прототипа стабильное боевое решение и определить, когда его нужно дорабатывать



Мониторинг модельного риска – необходимая часть работы модели в продуктиве

Что такое модельный риск?

- Модельный риск
 - ✓ Сумма потерь от принятия ошибочных решений при работе модели в продуктиве
- Меняющиеся условия реального мира вызывают деградацию модели
 - ✓ Проблемы в поставке данных
 - ✓ Данные на вход и факторы модели изменяются во времени (дрифт данных)
 - ✓ Закономерности в реальном мире меняются (дрифт метрик качества)
 - ✓ Модель применяется для сложных кейсов/новых данных (аномалии в прогнозах)
 - ✓ Обратная зависимость: модель влияет на реальный мир и на дальнейшее обучение модели (feedback loop)

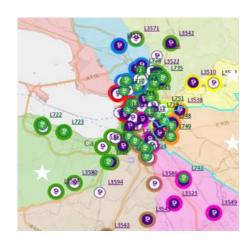
Доп. материалы: Model Performance Monitoring в компании Лента



Какие ключевые этапы моделирования должны быть реализованы для получения качественного ML решения?

Схема системы по работе с гео-локациями

Геопортал



Тул для использования бизнесом: управление процессом выбора и открытия локаций

MT

Интеграция и синхронизация данных и запросов geo-receiver

Взаимодействие с Геопорталом и управление цепочкой расчета

Big Data

geo feature service

Систематизация и автоматизация генерации гео-признаков

geo inference service

Инференс геомодели и составление отчета с интерпретацией прогнозов

feature store

- устранение разрыва dev-prod в данных и real time prod и обучения - ETL процесса расчета фич

model monitoring

Управление модельным риском – собственные сервисы для мониторинга офлайн и онлайн метрик



Ключевые роли в Big Data Lenta и их ответственность

Отдел управления продуктами

Product Owner

Максимизация ценности продукта для бизнеса, единая точка входа для бизнес-заказчика

Business Analyst

Анализ и описание бизнес-требований, подготовка простой аналитики и презентаций

Отдел разработки

Software Engineer

Разработка сервисов/приложений для бизнеса на основе решений продвинутой аналитики

DevOps

Организация эффективного создания и обновления решений продвинутой аналитики, контроль и мониторинг IT-инфраструктуры

System Analyst

Трансляция бизнес-требований в технические требования (ТЗ) для разработки

Отдел управления данными и IT

Data Engineer

Подготовка витрин, управление данными и базами данных

Data Analyst

Разработка и реализация аналитических (BI) решений

Отдел data science

Data Scientist

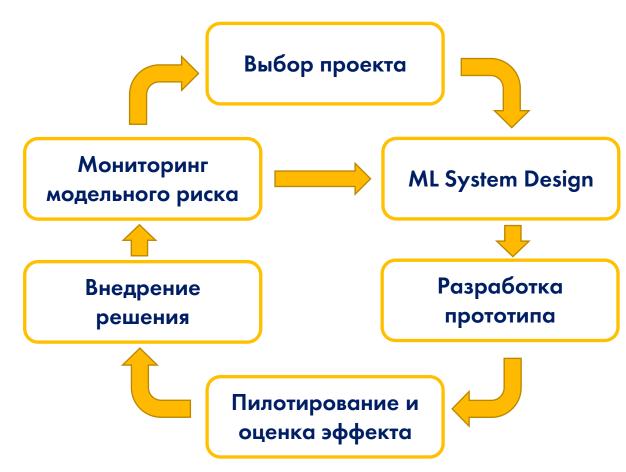
Разработка и реализация моделей продвинутой аналитики для решения задачи

ML Engineer Управление продуктивизацией моделей продвинутой аналитики, контроль инфраструктуры DS

Telegram-канал Reliable ML

Что делать, чтобы результат работы Big Data был применим в бизнес-процессах и приносил финансовую пользу







i.o.goloshchapova@gmail.com



@irina_goloshchapova





Ирина Голощапова Head of Data Science





Проблематика бизнес-процесса оценки пилотов



Многоэтапность и вовлечение большого количества сторон в процесс



Риск некорректной оценки результатов пилотов



Риск некорректного расчета финансового эффекта от ролл-аута



Задержки в проведении пилотов

Почему нужна единая математическая методика оценки пилотов?

Большинство оценок пилотов без использования единой методики не являются корректными

Причины на каждом этапе



Случайные расхождения пилотной и контрольной групп ТК:

- Высокая вероятность сказать, что эффект есть, когда его нет ошибка І-го рода
- Высокая вероятность сказать, что эффекта нет, когда он есть ошибка II-го рода

Некорректный вывод о значимости результата и возможности его экстраполяции

Причины на каждом этапе



Дизайн пилота: подбор пилотной группы ТК и

длительности пилота



Этап II Подбор контрольной группы ТК



Этап III Расчет эффекта



Этап IV результата



Пилотных ТК много/мало Длительность пилота критически низкая/высокая

Контрольные ТК выбраны без препилотного периода Не проведен А/А тест при



Алгоритм расчета не учитывает 😃 волатильность данных, сезонность, внешние факторы



Для вывода о результатах пилота используется точечная оценка



Не предусмотрены стат. тесты для сравнения групп



Некорректно считается доверительный интервал



ПОЧЕМУ НУЖНА ЕДИНАЯ МАТЕМАТИЧЕСКАЯ МЕТОДИКА ДЛЯ ПИЛОТОВ

На чем должна быть основана единая методика

- Случайные расхождения под контролем:
- Вероятность сказать, что эффект есть, когда его нет ошибка І-го рода ниже 15%
- Вероятность сказать, что эффекта нет, когда он есть ошибка II-го рода ниже 15 %

Решение о значимости результатов пилота и возможности его экстраполяции осуществляется на основе доверительного интервала эффекта

Этап II

Этап I Дизайн пилота: подбор пилотной группы ТК и длительности пилота

Пилотные ТК (число и конкретные магазины) и длительность выбираются с учетом снижения ошибок 1-2-го рода



Решения на каждом этапе

Контрольные ТК и алгоритм их выбора фиксируются с учетом снижения ошибок 1-2-го рода



Этап III Расчет эффекта

- Алгоритм расчета учитывает цель снижения ошибок 1-2-го рода, сезонность, критические внешние факторы
- ФакторыИспользуются стат. тесты

Решения на каждом этапе

Этап IV



Для вывода о результатах пилота используются точечная оценка и доверительный интервал с учетом формы распределения эффекта