



Фреймворк Reliable ML: оптимизация выбора гео- локаций открываемых магазинов

Ирина Голощапова
Head of Data Science

Обо мне

- ▶ Head of Data Science at LENTA
- ▶ Consultant at Skolkovo School of Management
- ▶ Lecturer at Sirius
- ▶ PhD in Economics
- ▶ In DS since 2009:
 - ▶ Statistical analysis and modelling at CMASF
 - ▶ DS in risks, macroeconomics and business process optimization at X5 Retail Group
 - ▶ DS Lead at Manchester University for social-impact projects
- ▶ Telegram-channel Reliable ML:
https://t.me/reliable_ml



@irina_goloshchapova



i.o.goloshchapova@gmail.com



План презентации

1

Что такое Reliable ML?

2

Reliable ML для оптимизации выбора гео-локаций открываемых магазинов

- Какова цель решения задачи с помощью ML?
- Какое место инициатива занимает в общей карте бизнес-процессов компании?
- Как ML-модель встроится в текущий бизнес-процесс инициативы, чтобы приносить пользу?
- Какие ключевые этапы моделирования должны быть реализованы для получения качественного ML решения?
- Как – с учетом бизнес-ограничений - мы можем оценить эффективность работы нашей модели?



План презентации



Что такое Reliable ML?



Reliable ML для оптимизации выбора гео-локаций открываемых магазинов

- Какова цель решения задачи с помощью ML?
- Какое место инициатива занимает в общей карте бизнес-процессов компании?
- Как ML-модель встроится в текущий бизнес-процесс инициативы, чтобы приносить пользу?
- Какие ключевые этапы моделирования должны быть реализованы для получения качественного ML решения?
- Как – с учетом бизнес-ограничений - мы можем оценить эффективность работы нашей модели?



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



План презентации

1

Что такое Reliable ML?

2

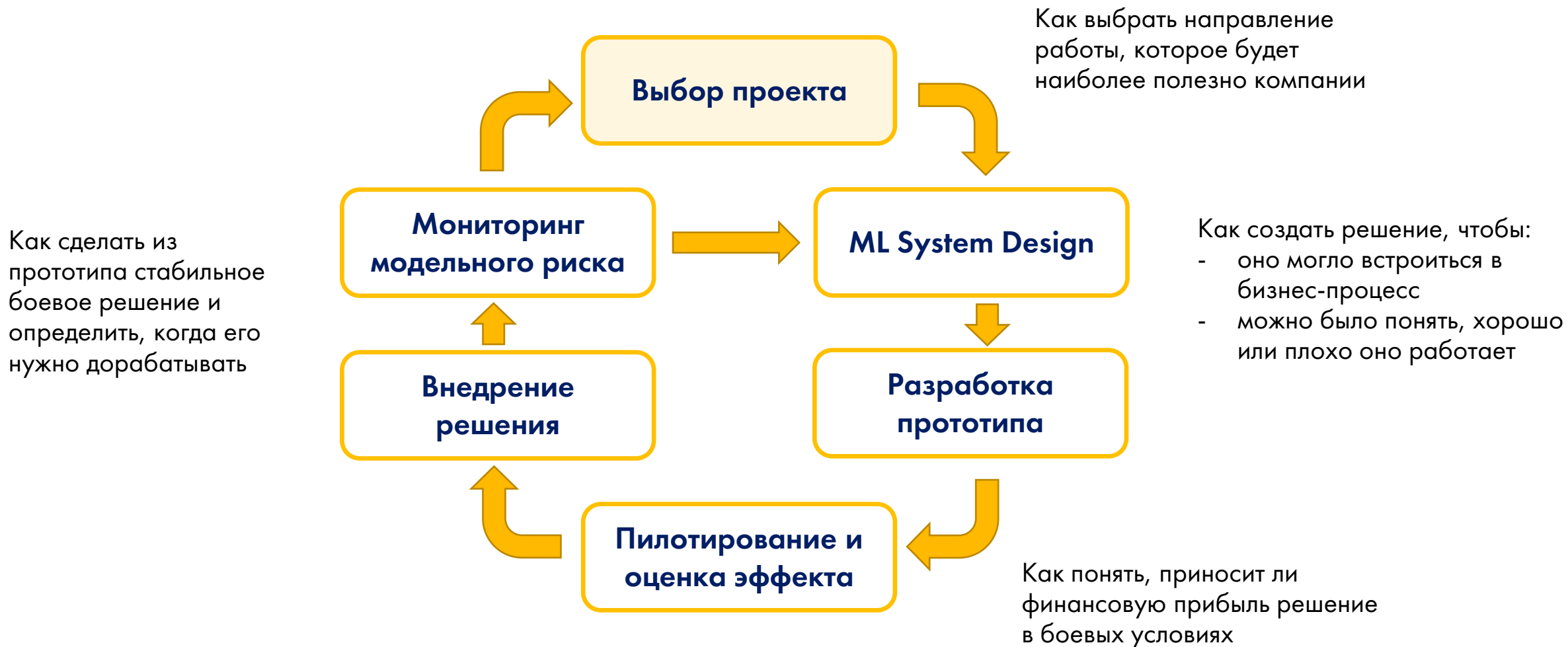
Reliable ML для оптимизации выбора гео-локаций открываемых магазинов

- Какова цель решения задачи с помощью ML?
- Какое место инициатива занимает в общей карте бизнес-процессов компании?
- Как ML-модель встроится в текущий бизнес-процесс инициативы, чтобы приносить пользу?
- Какие ключевые этапы моделирования должны быть реализованы для получения качественного ML решения?
- Как – с учетом бизнес-ограничений - мы можем оценить эффективность работы нашей модели?



Какова цель решения задачи с помощью ML?

Фреймворк Reliable ML



Какова цель решения задачи с помощью ML?

Будем решать задачу оптимизации выбора гео-локаций открываемых магазинов



Прогнозировать выручку открываемых магазинов

Показывать прогноз в 1й и 2й год работы нового магазина с учетом его внутренних параметров (площадь, дата открытия, формат, и др.) и внешних характеристик локации таким образом, чтобы **максимизировать финансовую выгоду** от органического роста компании

Какова цель решения задачи с помощью ML?

Погрузимся детальнее в постановку задачи

Исторический выбор локаций

- **Низкая точность прогноза** – прогноз основан на экспертном знании рынка и не учитывает исторические данные
- **Неглубокая аналитика** – выводы об ожидаемой выручке магазина опираются на крайне ограниченное количество данных
- **Сложность и низкая скорость управления** – ручной процесс расчета ожидаемой выручки ограничен по скорости



Целевая модель

- **Высокая точность прогноза** – прогноз основан на исторических закономерностях
- **Проактивный подход** – выводы опираются на массив всех известных исторических данных и множество факторов
- **Упрощение управления** – рост скорости принятия решений и выявления оптимальных локаций для открытия новых точек продаж

План презентации

1

Что такое Reliable ML?

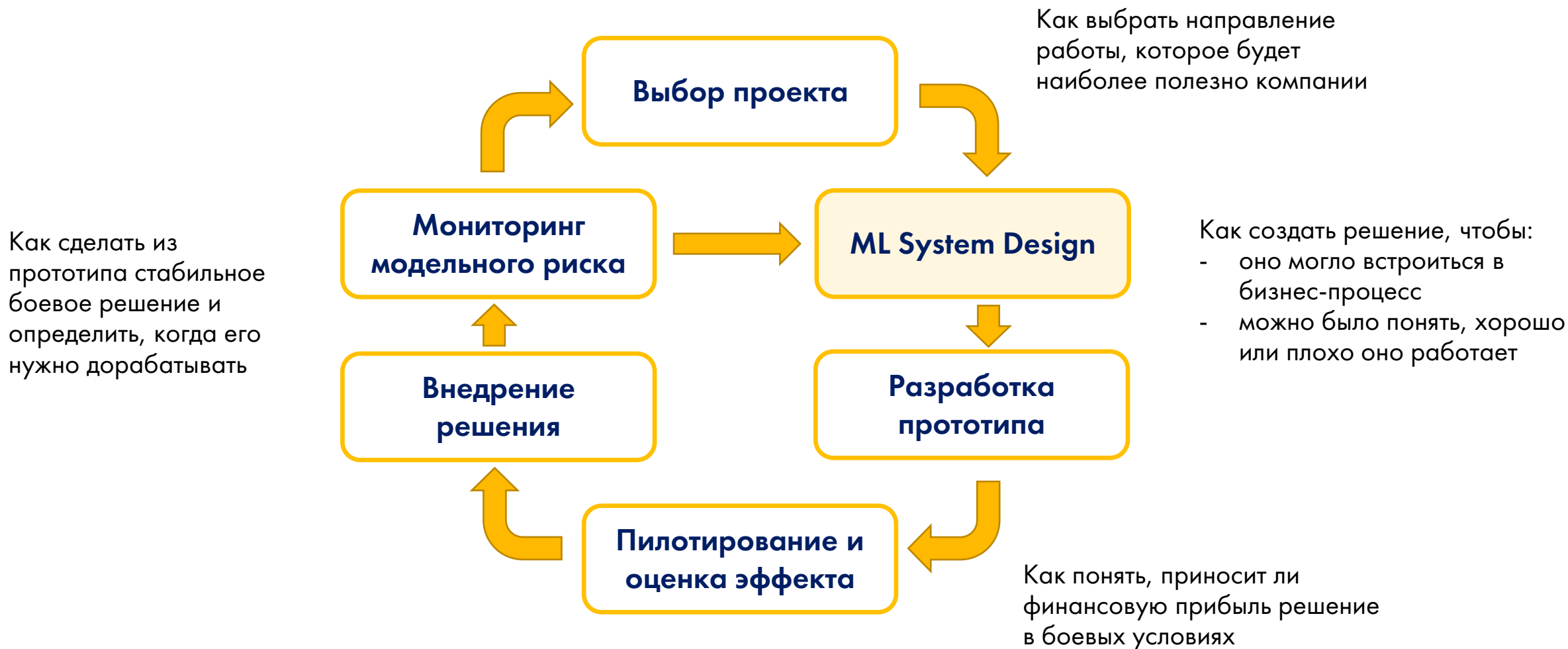
2

Reliable ML для оптимизации выбора гео-локаций открываемых магазинов

- Какова цель решения задачи с помощью ML?
- Какое место инициатива занимает в общей карте бизнес-процессов компании?
- Как ML-модель встроится в текущий бизнес-процесс инициативы, чтобы приносить пользу?
- Какие ключевые этапы моделирования должны быть реализованы для получения качественного ML решения?
- Как – с учетом бизнес-ограничений - мы можем оценить эффективность работы нашей модели?

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

Фреймворк Reliable ML



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

Выбор гео-локации и прогноз выручки – часть управления органическим ростом компании



Что для нас важно?

- Прогноз ожидаемой выручки в центре множества процессов и должен удовлетворять их потребностям
- Модель прогноза выручки в компании не одна
- Финансовая выгода зависит не только от выручки, но и от ожидаемых и фактических издержек

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

Прогноз выручки для конкретных локаций – основа бизнес-процесса

Модель должна учитывать ключевые факторы, способные повлиять на финансовую выгоду от открытия магазина в кратко- и долгосрочном периодах

Прогноз выручки в открываемом ТК



План презентации

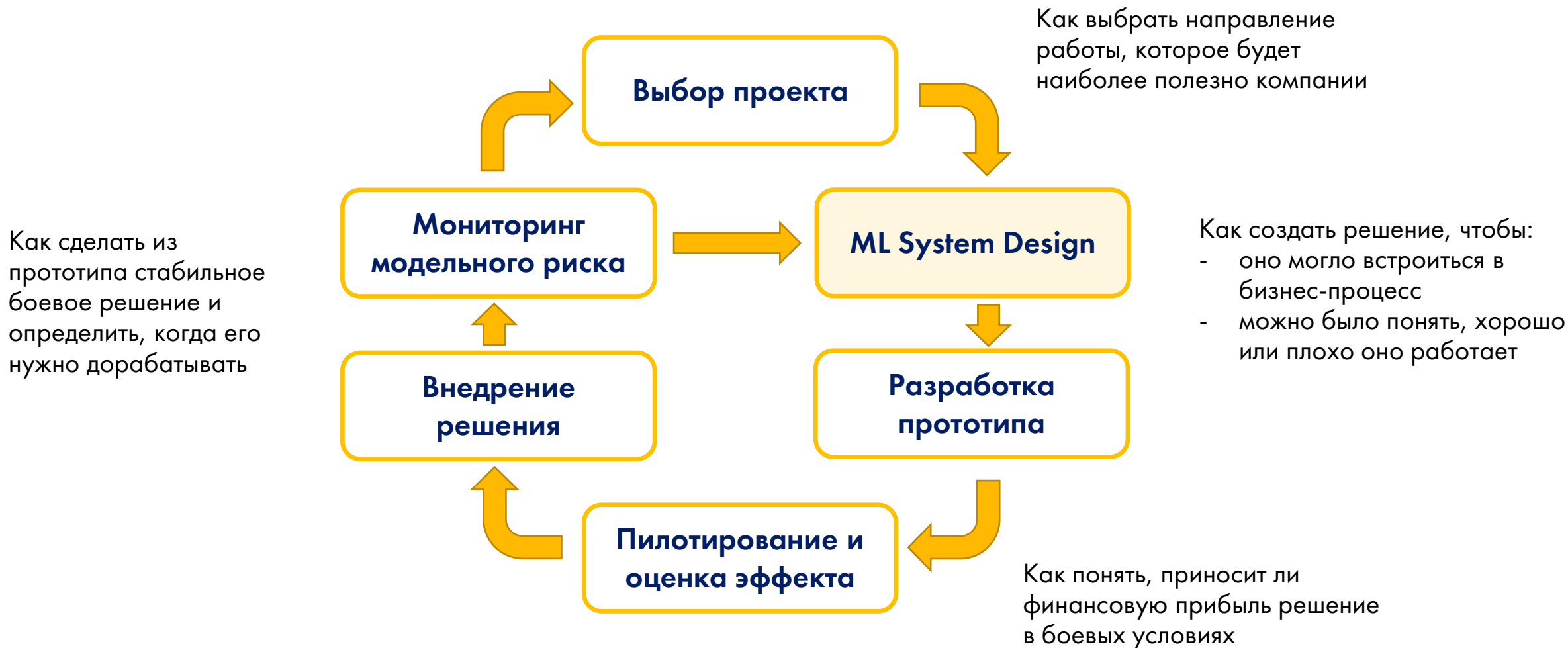
1 Что такое Reliable ML?

2 Reliable ML для повышения эффективности промо в Ленте

- Какова цель решения задачи с помощью ML?
- Какое место инициатива занимает в общей карте бизнес-процессов компании?
- Как ML-модель встроится в текущий бизнес-процесс инициативы, чтобы приносить пользу?
- Какие ключевые этапы моделирования должны быть реализованы для получения качественного ML решения?
- Как – с учетом бизнес-ограничений - мы можем оценить эффективность работы нашей модели?

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

Фреймворк Reliable ML



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

Изменение параметров промо-активностей

Как Data Scientists представляют себе процесс открытия магазина в продуктовом ритейле в начале проекта?

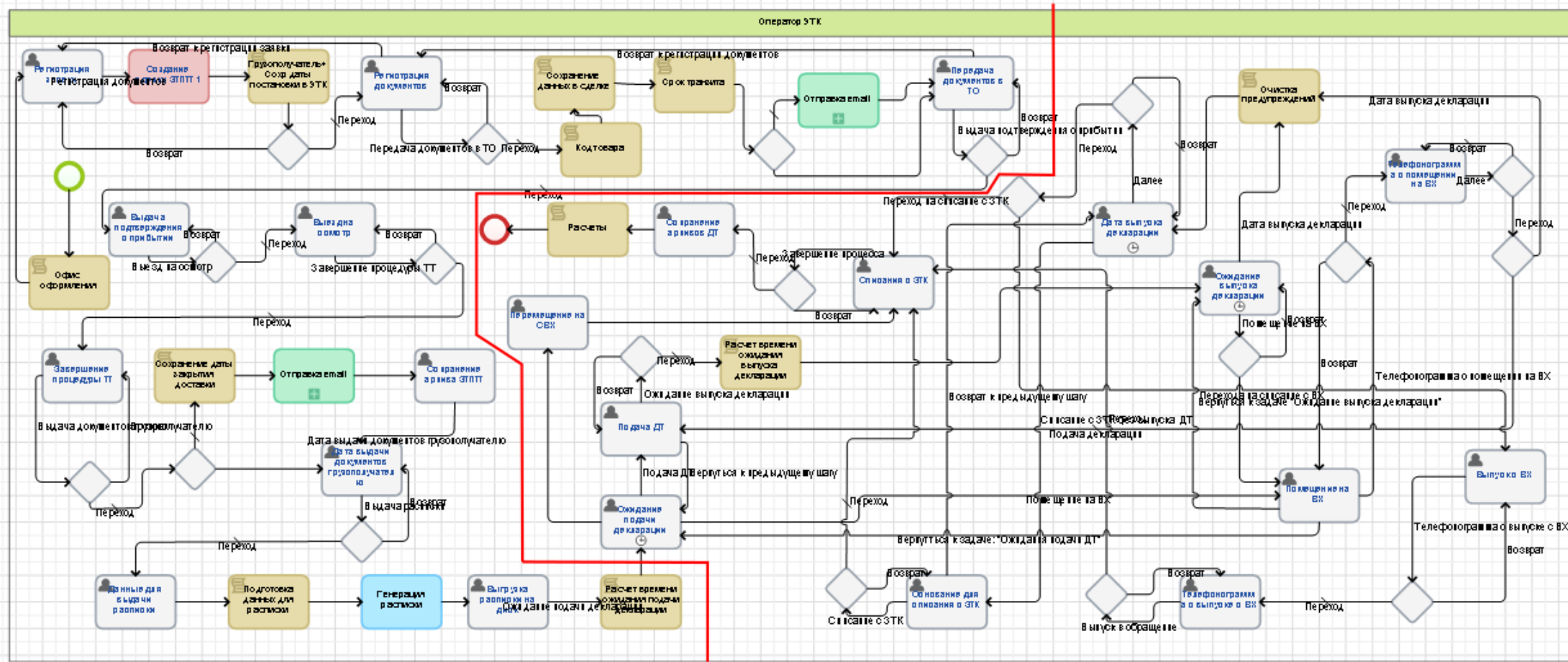
Рекомендация DS модели
по открытию магазина



Открытый
магазин,
счастливая
Лента и
покупатели

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

Как в продуктовом ритейле в реальности открывают магазины?



* Абстрактный пример сложного бизнес-процесса



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
- **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

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

Виды продвинутой аналитики



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

Последовательное развитие продукта промо для получения максимального коммерческого эффекта



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

Последовательное развитие продукта для получения максимального коммерческого эффекта

Модель должна учитывать ключевые факторы, способные повлиять на финансовую выгоду от открытия магазина в кратком и долгосрочном периодах

Прогноз выручки в открываемом ТК



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

Последовательное развитие продукта для получения максимального коммерческого эффекта

Модель должна учитывать ключевые факторы, способные повлиять на финансовую выгоду от открытия магазина в кратком и долгосрочном периодах

Прогноз выручки в открываемом ТК



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

Последовательное развитие продукта промо для получения максимального коммерческого эффекта

Модель должна учитывать ключевые факторы, способные повлиять на финансовую выгоду от открытия магазина в кратком и долгосрочном периодах

Прогноз выручки в открываемом ТК



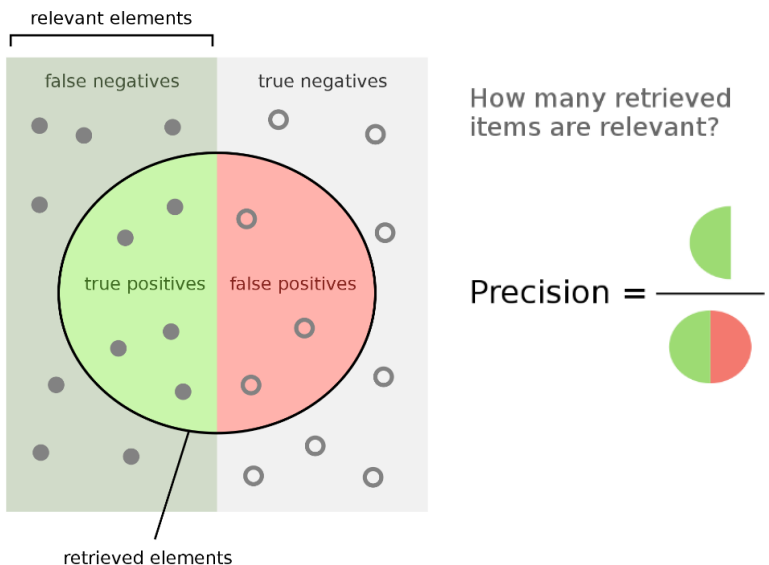
Как ML-модель встроится в текущий бизнес-процесс инициативы, чтобы приносить пользу?

Метрики качества решения должны отражать бизнес-результат

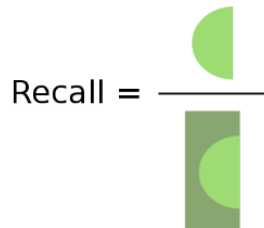
Технические метрики

Качество прогноза выручки

- Средневзвешенная средняя процентная ошибка (WAPE)
- Средняя процентная ошибка (MAE)



How many relevant items are retrieved?



Бизнес-метрики

Применимость модели для бизнес процесса

Общее качество рекомендаций:

- **Total Accuracy:** число успешных рекомендаций / число открытых магазинов
- **Positive Hit Rate (Recall)** – число успешных рекомендаций / число успешных магазинов
- **Negative Hit Rate** – число рекомендаций не открывать магазин / число неуспешных магазинов
- **Precision** – число успешных рекомендаций / число рекомендаций магазинов к открытию

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

Успех выбора гео-локации зависит не только от ожидаемой выручки, но и от ожидаемых затрат

Цель модели – итоговая финансовая выгода от открытий магазинов: успешные открытия и предотвращение открытий неуспешных магазинов.

	Store Status	Store Success*	Model reco**	Effect calculation**	Loss or Profit
1	Opened	No	Not to open	1	Savings
2	Opened	No	To open	0	Loss
3	Opened	Yes	Not to open	1	Loss
4	Opened	Yes	To open	0	Income

* Может быть рассчитан через пороговое значение выручки, исходя из требований к финансовым результатам магазина (NPV, EBITDA, DPP, IRR)

** Может быть рассчитан на основе EBITDA Annualised, Initial Investments, NPV



План презентации

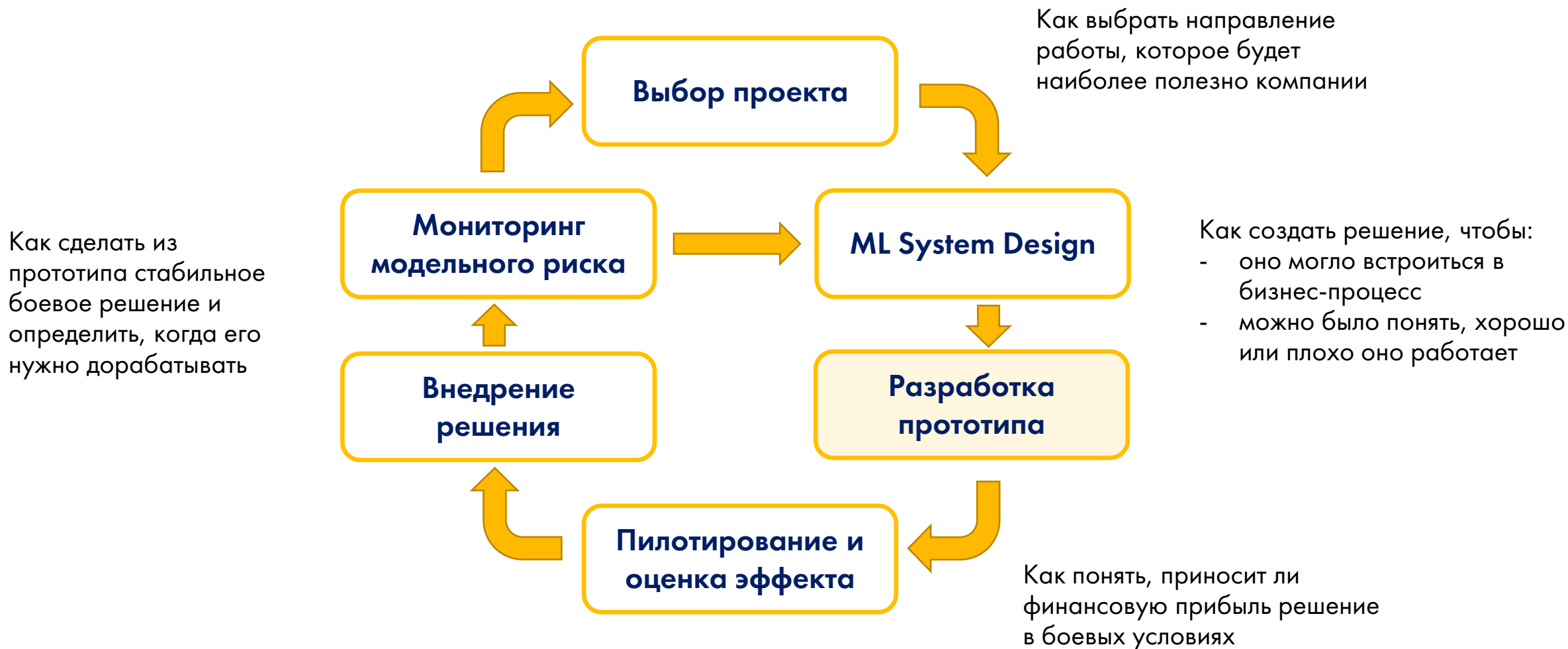
1 Что такое Reliable ML?

2 **Reliable ML для повышения эффективности промо в Ленте**

- Какова цель решения задачи с помощью ML?
- Какое место инициатива занимает в общей карте бизнес-процессов компании?
- Как ML-модель встроится в текущий бизнес-процесс инициативы, чтобы приносить пользу?
- **Какие ключевые этапы моделирования должны быть реализованы для получения качественного ML решения?**
- Как – с учетом бизнес-ограничений - мы можем оценить эффективность работы нашей модели?

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

Фреймворк Reliable ML



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

Ключевые этапы для реализации оптимизации выбора гео-локаций открываемых магазинов



Найти вовлеченную и компетентную команду

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

Ключевые этапы для ML System Design и разработки прототипа

Имплементация и оценка эффекта

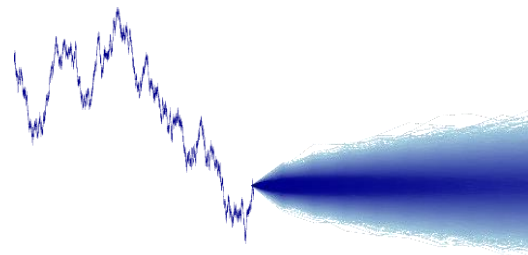


I. Подготовка и препроцессинг данных

- Расчет множества гео-факторов в различных кечментах
- Учет историчности данных
- Проверка корректности данных

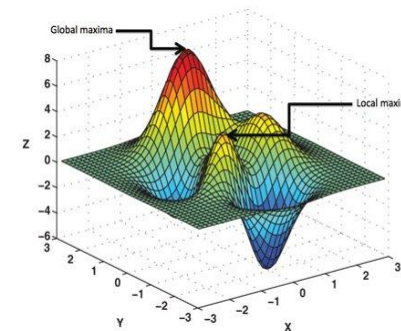
II. Моделирование

- Обучение моделей прогноза выручки
- Выбор оптимального алгоритма построения модели на основе технических метрик качества



III. Оптимизация

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



IV. Бизнес-проверка

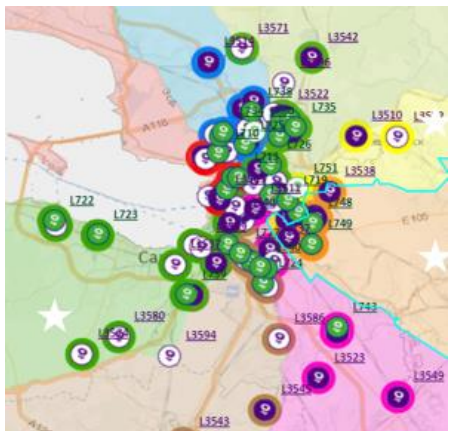
- Бизнес-анализ прогнозов, учет ОС
- Планирование доработок и новых исследований с учетом обратной связи от экспертов



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

ИТ

Интеграция и синхронизация данных и запросов

Big Data

geo feature service

Систематизация и автоматизация генерации гео-признаков

feature store

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

geo-receiver

Взаимодействие с Геопорталом и управление цепочкой расчета

geo inference service

Инференс геомодели и составление отчета с интерпретацией прогнозов

model monitoring

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

План презентации

1

Что такое Reliable ML?

2

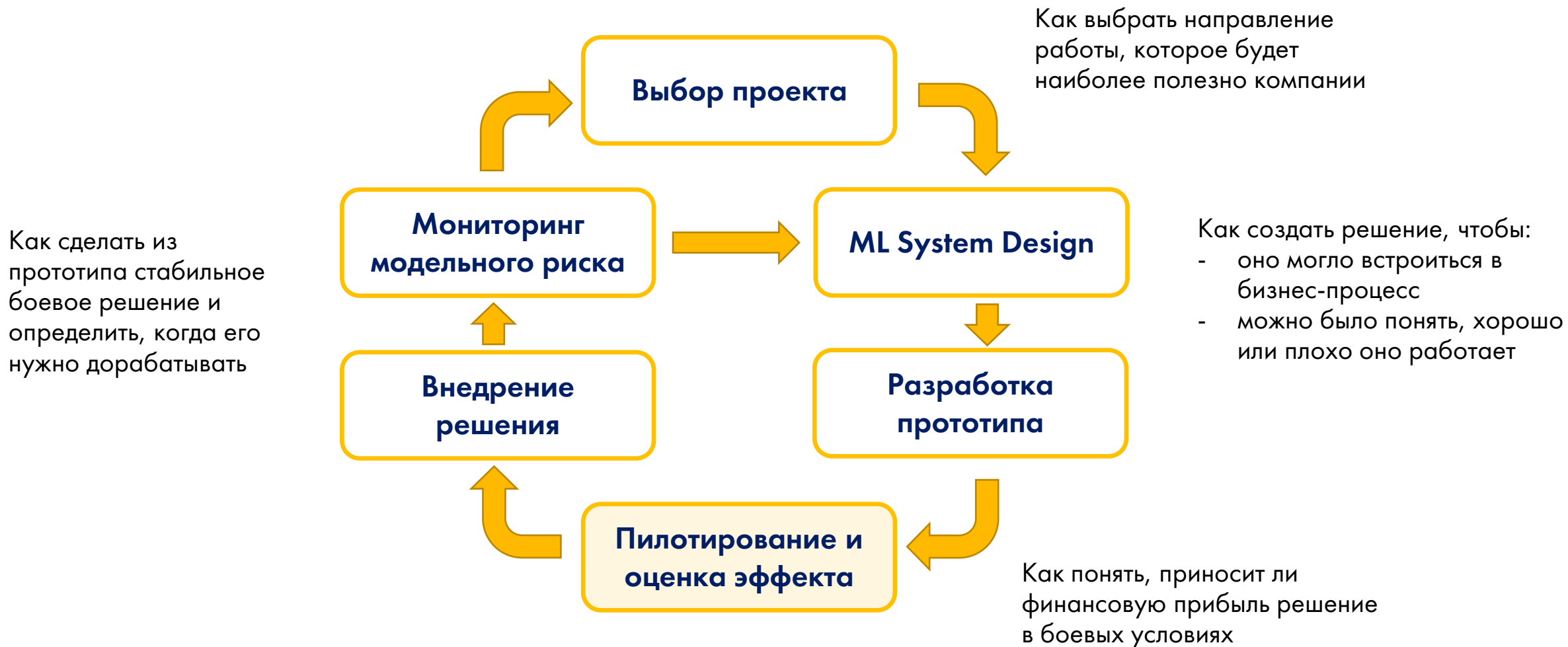
Reliable ML для повышения эффективности промо в Ленте

- Какова цель решения задачи с помощью ML?
- Какое место инициатива занимает в общей карте бизнес-процессов компании?
- Как ML-модель встроится в текущий бизнес-процесс инициативы, чтобы приносить пользу?
- Какие ключевые этапы моделирования должны быть реализованы для получения качественного ML решения?
- Как – с учетом бизнес-ограничений - мы можем оценить эффективность работы нашей модели?



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

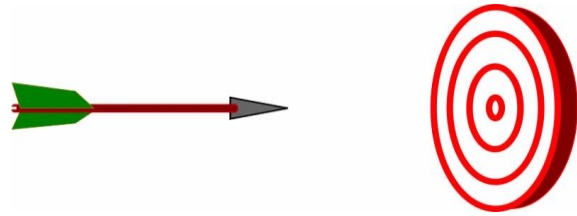
Фреймворк Reliable ML



Дс Как – 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**



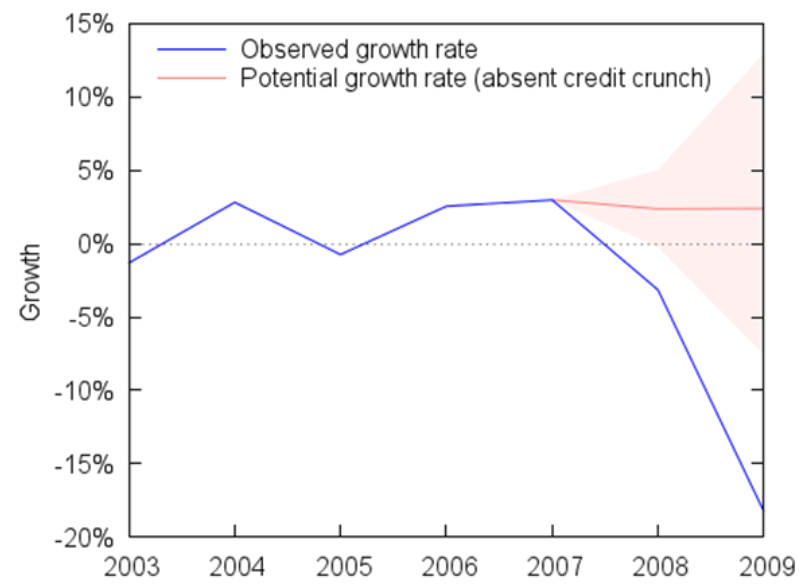
- ▶ Простое сравнение средних не работает
- ▶ Не можем отличить эффект от
 - случайности
 - влияния внешних факторов
 - специфики выбранных в пилот и контроль объектов
 - манипуляций
- ▶ Не знаем, повторится ли найденный эффект при новом эксперименте
- ▶ **Делаем некорректный вывод о ролл-ауте инициативы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

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

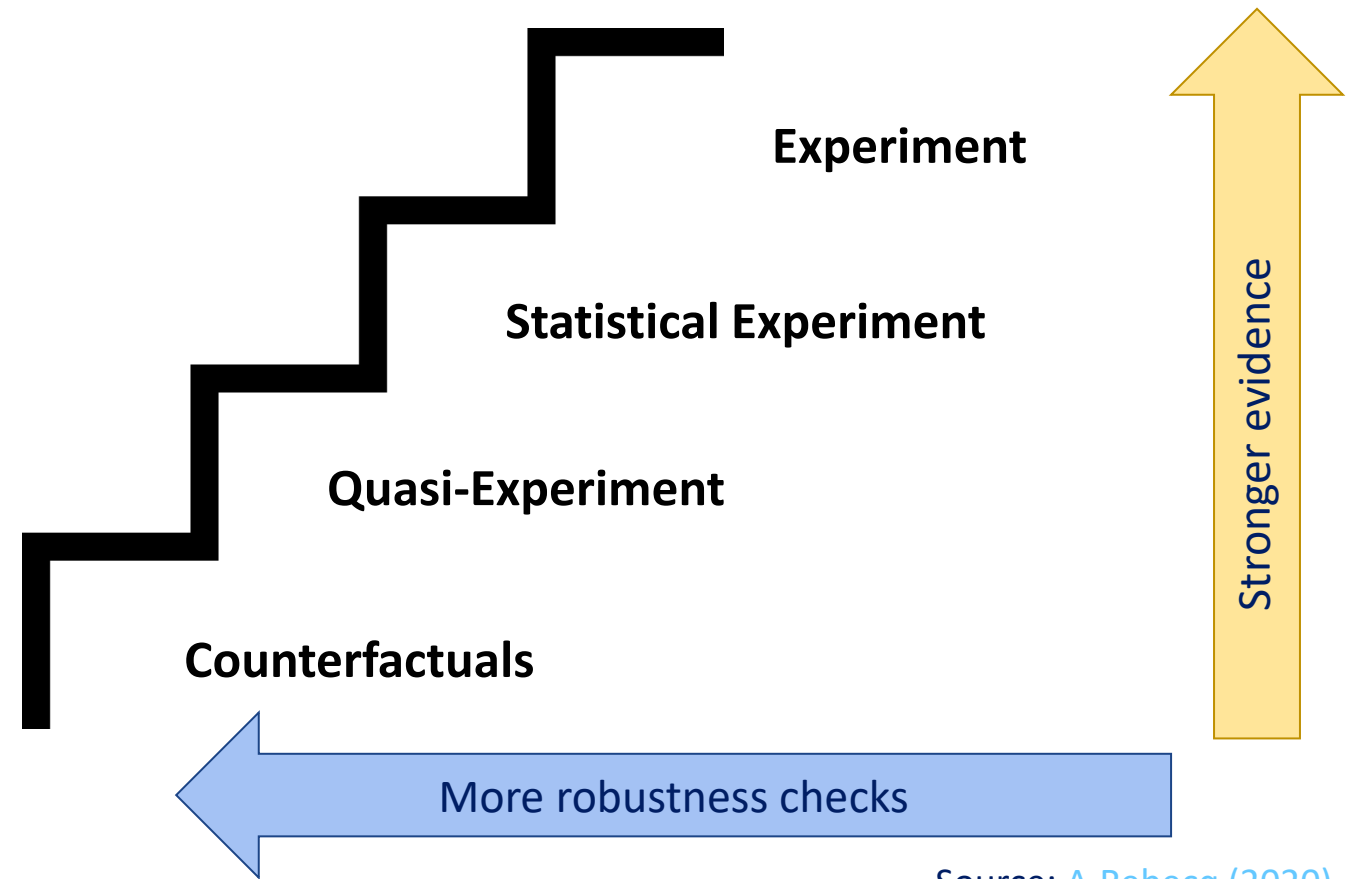
Что делать, если АБ-тестирование не работает

Решение

- Необходимо использовать группу методов причинно-следственного анализа – Counterfactual Analysis



Evidence Ladder



Source: [A.Rebecq \(2020\)](#)



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

Что делать в случае оптимизации выбора гео-локации открываемых магазинов

Особенности оценки эффекта

- Пилоты, как правило, проводятся на отдельных категориях продаж:
 - Нет возможности получить группы А и Б – можем только варьировать сам бизнес-процесс, усиливая роль одной или другой модели
 - Есть параметры, на которые мы не влияем ML-моделью (ожидаемые издержки)
 - Всегда игра с неполной информацией (не можем посчитать метрики по неоткрытым магазинам)

Что делать

- Агрегированный мониторинг изменения метрик во времени
- Бенчмаркинг рыночных решений



Telegram-канал Reliable ML

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



**Спасибо
за внимание!**



**Ирина Голощапова
Head of Data Science**

