# Дизайн систем машинного обучения

## 10. Жизненный цикл модели

# План курса

1) ~~Практическое применение машинного обучения~~
2) ~~Основы проектирования ML-систем~~
3) ~~Обучающие данные~~
4) ~~Подготовка и отбор признаков~~
5) ~~Выбор модели, разработка и обучение модели~~
6) ~~Оценка качества модели~~
7) ~~Развертывание~~
8) ~~Диагностика ошибок и отказов ML-систем~~
9) ~~Мониторинг и обучение на потоковых данных~~
10) **Жизненный цикл модели — Вы находитесь здесь**
11) Отслеживание экспериментов и версионирование моделей
12) Сложные модели: временные ряды, модели над графами
13) Непредвзятость, безопасность, управление моделями
14) ML инфраструктура и платформы
15) Интеграция ML-систем в бизнес-процессы

# ГОСТ Р 57193-2016

4.1.19 жизненный цикл (life cycle): Развитие системы, продукции, услуги, проекта или другой создаваемой человеком сущности от замысла до списания.

4.1.20 модель жизненного цикла (life cycle model): Структурная основа процессов и действий, относящихся к жизненному циклу, которая также служит в качестве общего эталона для установления связей и понимания.

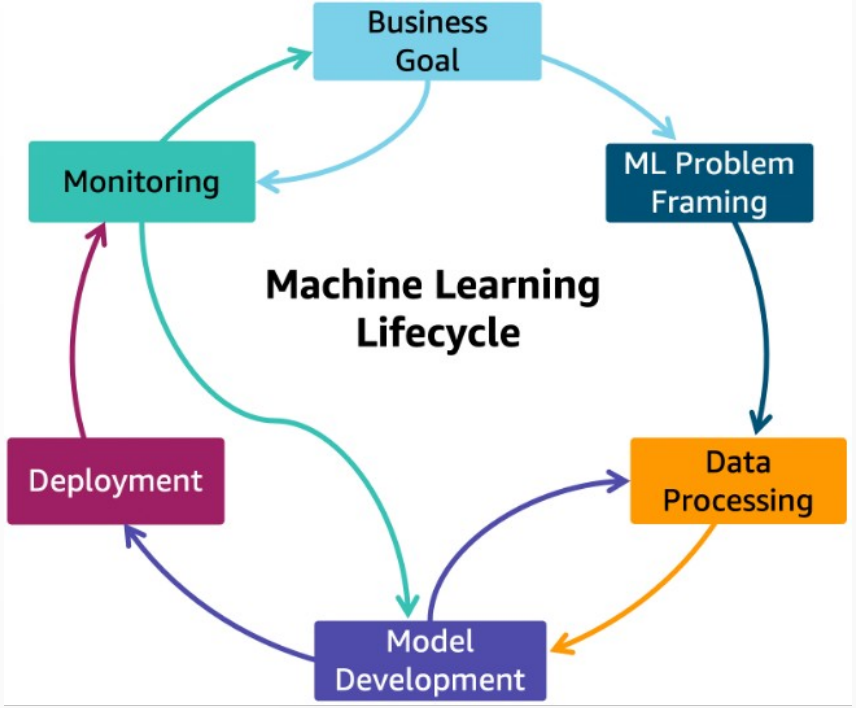
# CRISP-DM (1999)

- Модель жизненного цикла
- ~~Устарела~~
  Проверена временем
- ~~Нереалистичная~~
  Красивая
- ~~Не используется на практике~~
  Подходит для обучения студентов


https://ru.wikipedia.org/wiki/CRISP-DM

# AWS machine learning lifecycle

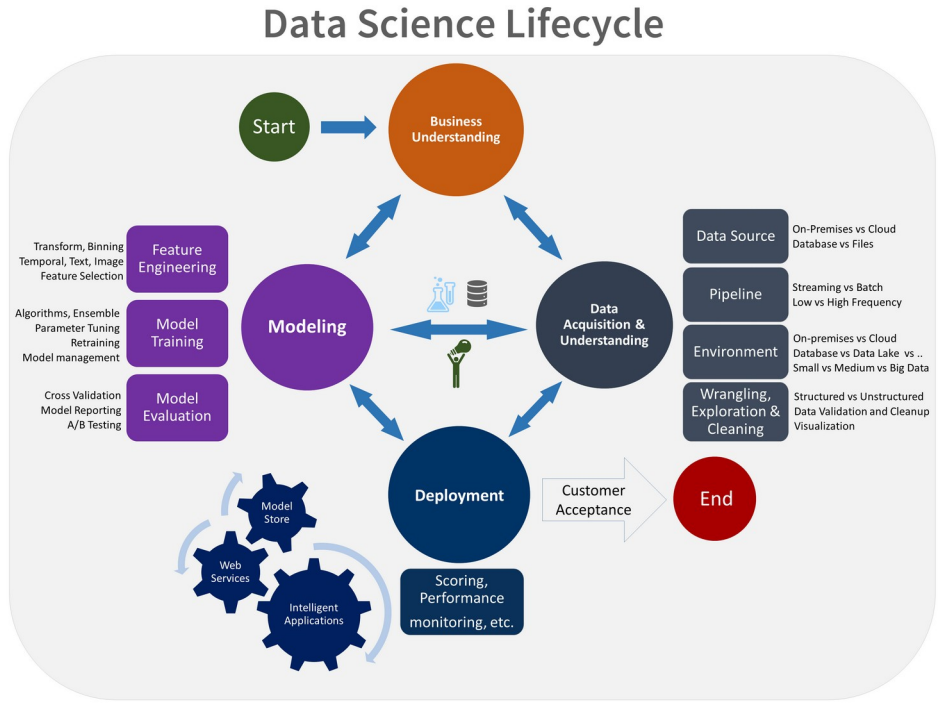
- Решаем
- Делаем
- Смотрим
- Дорабатываем
- …




https://docs.aws.amazon.com/wellarchitected/latest/machine-learning-lens/well-architected-machine-learning-lifecycle.html

# The Team Data Science Process

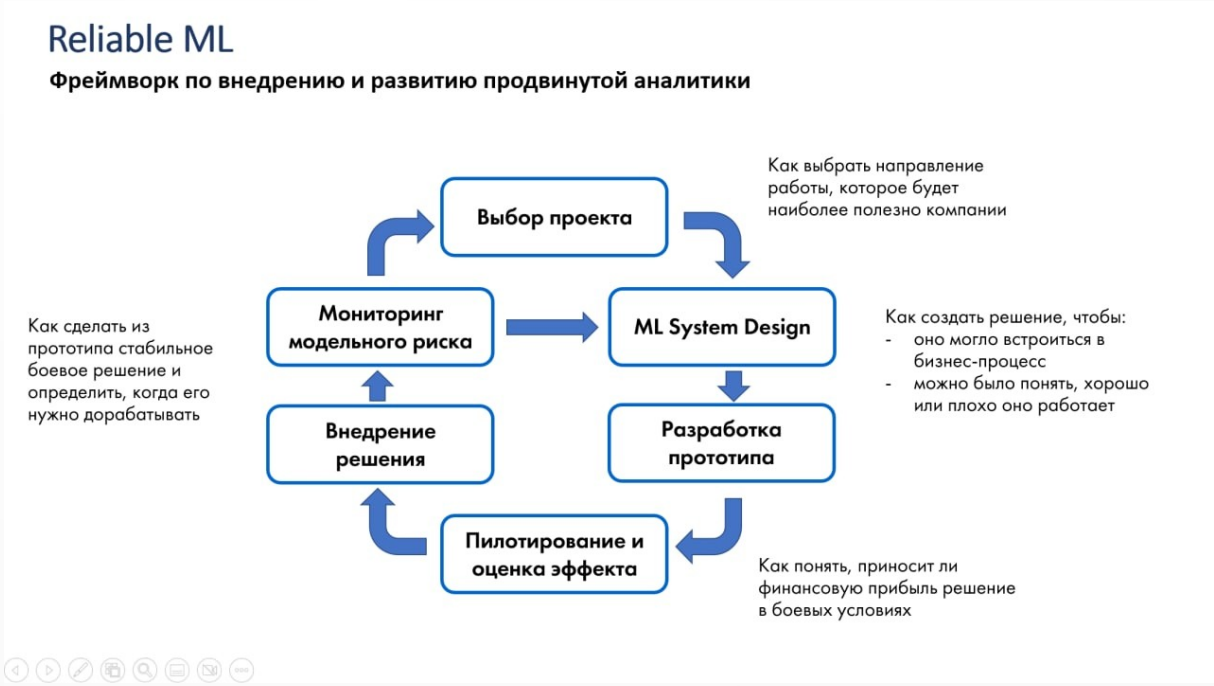
- Решаем
- Делаем
- Смотрим
- Дорабатываем
- …




https://learn.microsoft.com/en-us/azure/architecture/data-science-process/lifecycle

# Reliable ML

- Решаем
- Делаем
- Смотрим
- Дорабатываем
- …




#ReliableML

# Модель придется дорабатывать

- Сдвиг данных Data Shift
- Меняющиеся бизнес-требования
- Как часто переобучать модель?
- Достаточно ли переобучить модель на новых данных?
- Как сравнить производительность новой и старой модели?
- Как автоматизировать дообучение моделей?

https://medium.com/@nagasanjayvijayan/continuous-training-of-ml-models-7d8acaf44dda

# Новые данные или новая архитектура

- Акцент на данных:
  - Есть возможность переобучать модель чаще, и это дает отдачу
  - Есть возможность дочистить данные, и это дает отдачу
  - Есть возможность добавить еще данных, и это дает отдачу
- Акцент на новой архитектуре:
  - Все, что можно сделать с данными, уже сделали

# Постоянное дообучение модели

- Идея:
  - дообучать модель каждый раз, как приходит новый пример
- Проблемы:
  - Катастрофическое забывание →→
  - Большая вычислительная нагрузка
  - Архитектура оптимизирована для инференса, не для обучения
- Решение:
  - Дообучать батчами (раз в день или каждые 1000 точек, например)

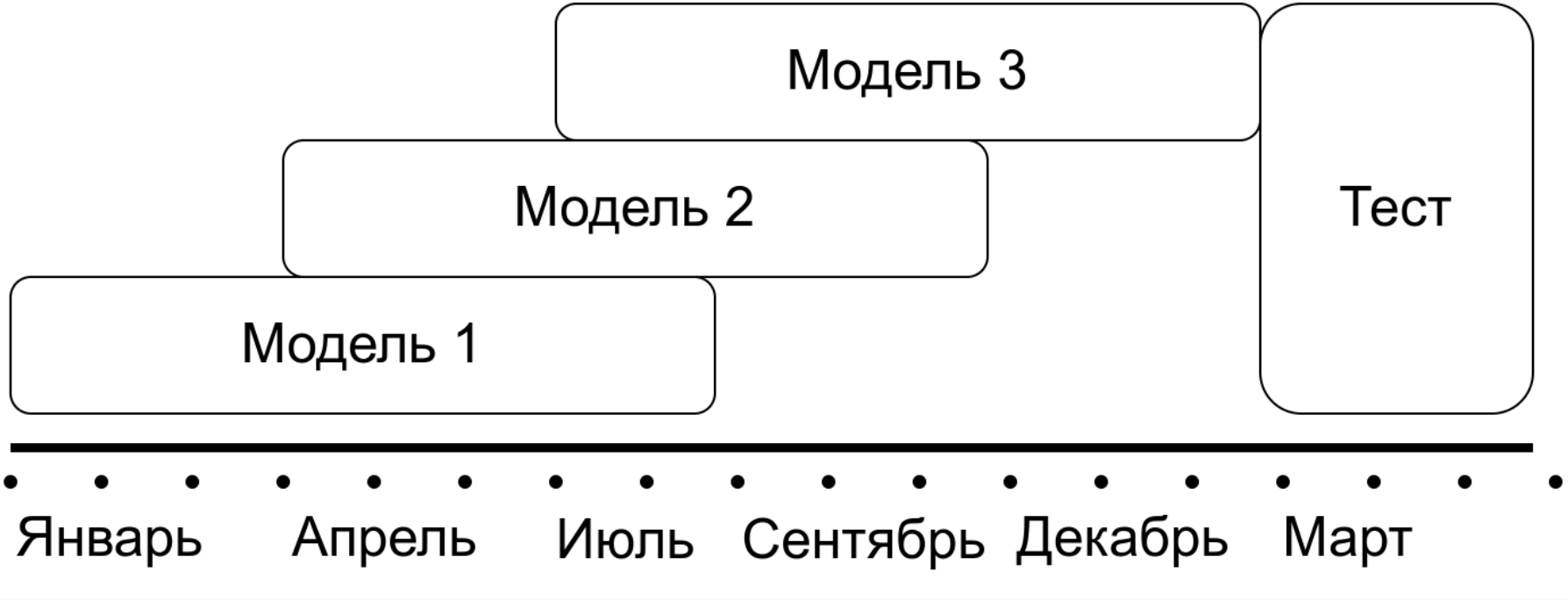
# Как часто дообучать модель?

- Много ли данных поступает?
- Быстро ли меняются распределения?
- Заметны ли улучшения после переобучения?
- В производстве — чаще, чем меняют основные датчики (> 2x)

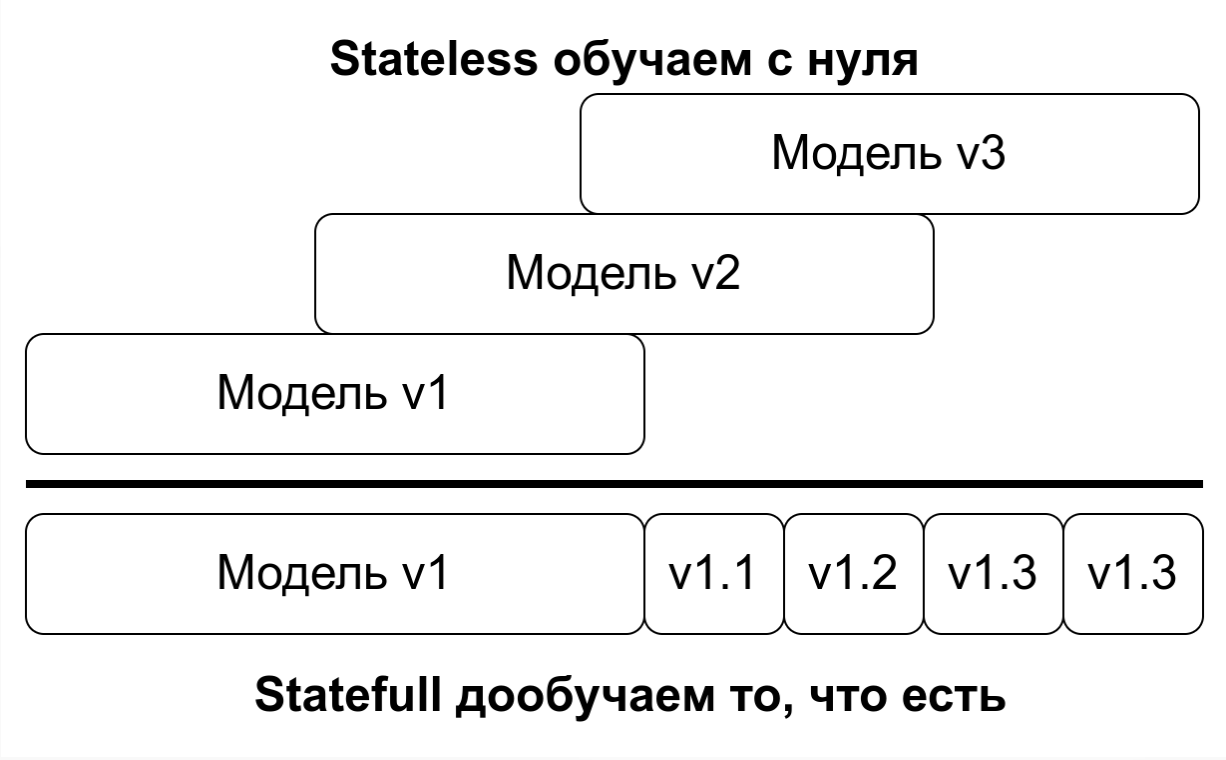
Prediction accuracy clearly degrades for both models as the delay between training and test set increases. For both models it can been seen that NE can be reduced by approximately 1% by going from training weekly to training daily.

https://research.fb.com/wp-content/uploads/2016/11/practical-lessons-from-predicting-clicks-on-ads-at-facebook.pdf

# Бэктесты


Модель 3
Модель 2
Модель 1
Тест
Январь
Апрель
Июль
Сентябрь
Декабрь
Март

# Переобучать или дообучать?

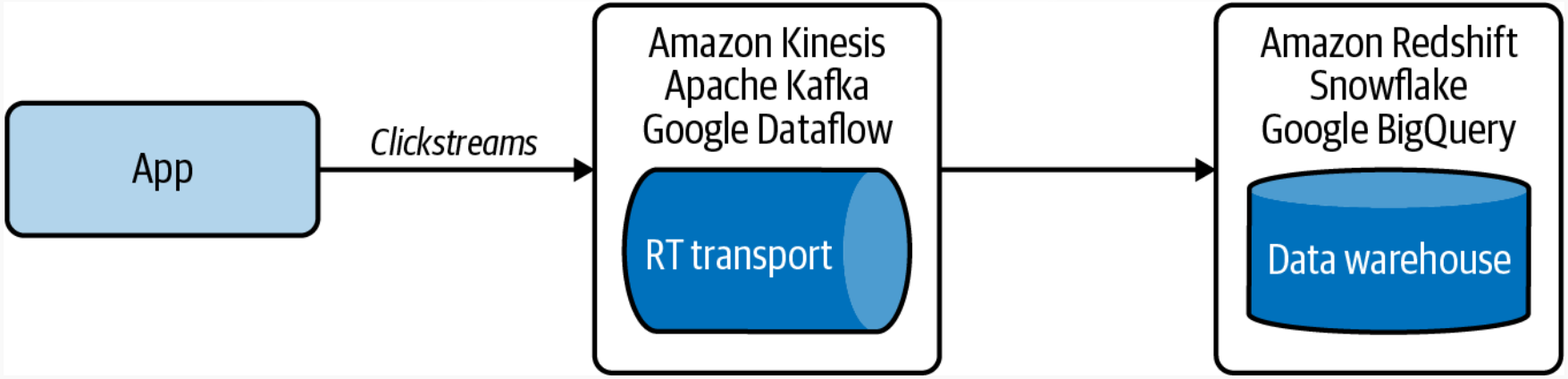


# Что ограничивает частоту обучения

- Проблемы с доступом к данным
- Проблемы с доступом к разметке
- Скорость переобучения
- Проблемы с оценкой и сравнением моделей
- Ограничения ML-алгоритмов

# Дообучение: доступ к свежим данным

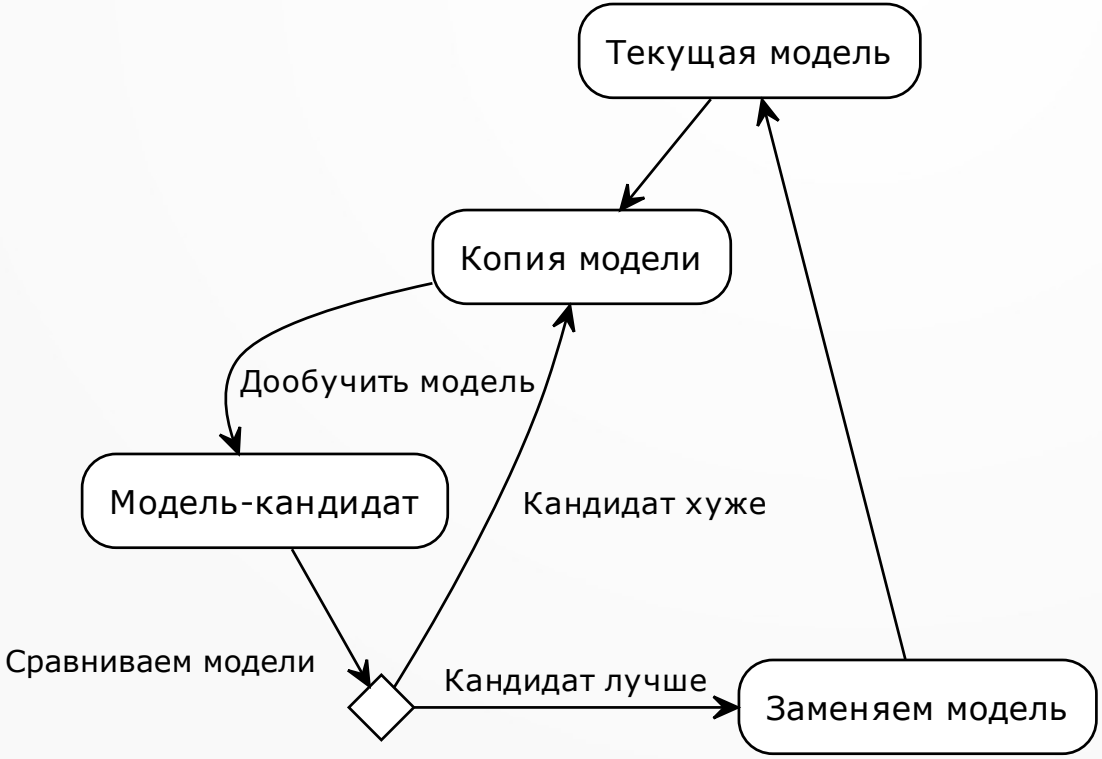
- Проблема:
  - Нужно часто выгружать данные из хранилища
- Решение:
  - Брать новые данные из конвеера данных. Кешировать признаки.



# Дообучение: доступ к свежей разметке

- Проблема:
  - Мы не можем учиться быстрее, чем приходит разметка
- Решение:
  - Естественные метки natural labels
  - Обратная связь от пользователей
  - Программная слабая разметка (например, Snorkel)

# Дообучение: время на оценку модели



# Дообучение: ограничения алгоритмов

- Небольшие нейронные сети дообучать просто
- Большие нейронные сети дообучать долго
- Ансамбли градиентного бустинга дообучать сложно
  - Достраиваем деревья
  - Время от времени учим с нуля или делаем дистилляцию
- Алгоритмы матричной факторизации дообучать сложно
  - Есть варианты, см recalculate_* в implicit

# Четыре оттенка онлайн-обучения

- Manual stateless retraining
  - Переобучение по запросу
- Automated stateless retraining
  - Переобучение по расписанию
- Automated statefull retraining
  - Управление версиями моделей и данных
- Continual learning
  - Гибкое расписание дообучения (время, качество, объем, сдвиг)

# Тестирование вживую

- Как оценить качество дообученной модели?
- Оцениваем качество на старых данных
  - Сдвиг данных искажает качество
- Бэктесты (ретротесты) →
  - Переобучение на тестовой выборке
- Тестировать на живых данных
  - Но как?

# Тестирование вживую

- Теневое развертывание Shadow Deployment
- Пробный релиз Canary Release
- А/Б тестирование
- Чередование Interleaving
- Многорукие бандиты

# Shadow Deployment

- Разворачиваем новую модель
- Дублируем на нее трафик
- Записываем предсказания, но не отдаем клиенту
- Анализируем, ошиблась или нет
- Проблемы:
  - Мы удваиваем затраты на инференс
  - Мы выкатываем новую модель с задержкой
  - Нужно учитывать  смещение выборки selection bias →

# Canary Release

- Разворачиваем новую модель
- Направляем на нее небольшую часть трафика
- Ловим грубые ошибки
- Перед А/Б тестом или вместо него, если изменение небольшое

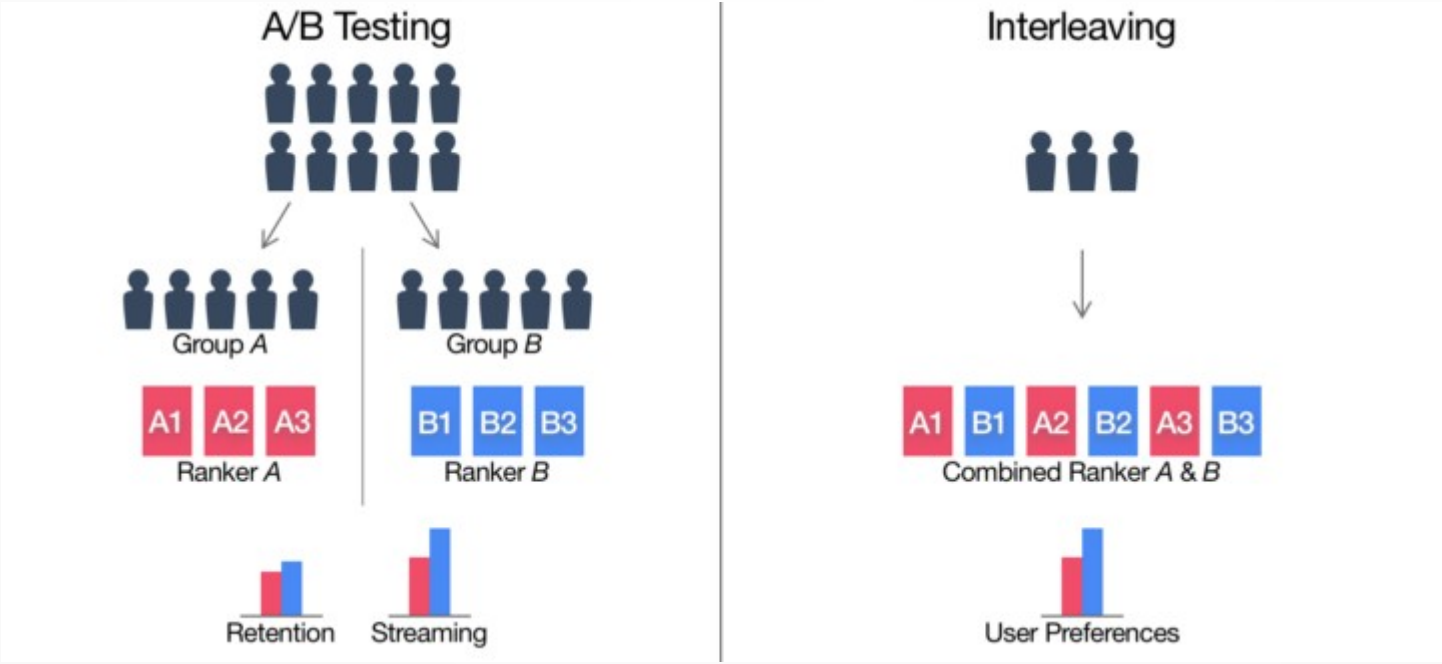
Automated Canary Analysis at Netflix with Kayenta

# A/B Testing

- Разворачиваем новую модель
- Направляем на нее часть трафика
- Сравниваем качество и отзывы пользователей
- A/A/B Test
- Проблемы:
  - Рандомизация трафика
  - Статистическая мощность (объем выборки)

https://www.algorithmicmarketingbook.com/
https://hbr.org/2017/09/the-surprising-power-of-online-experiments

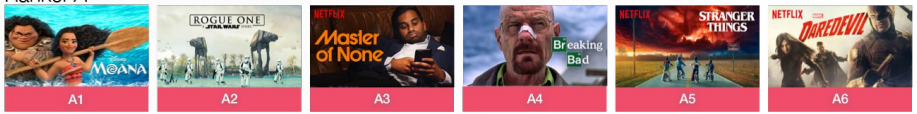
# Interleaving



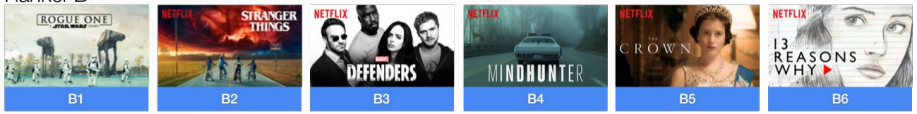

https://netflixtechblog.com/interleaving-in-online-experiments-at-netflix-a04ee392ec55

# Interleaving

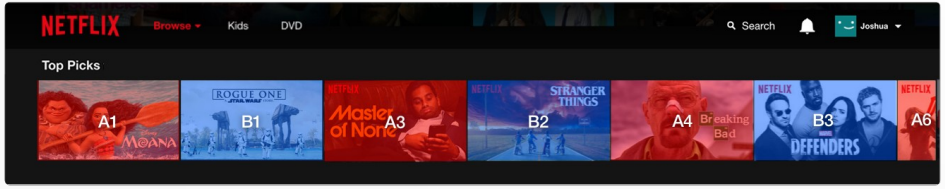
Ranker *A*



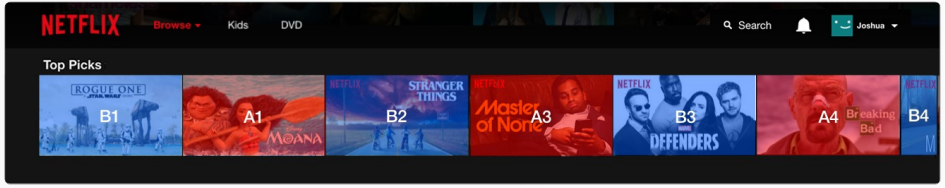

Ranker *B*




If Ranker *A* chooses first




If Ranker *B* chooses first




https://netflixtechblog.com/interleaving-in-online-experiments-at-netflix-a04ee392ec55

# Bandits

- Разворачиваем несколько моделей
- Случайно направляем на них трафик
- Чем лучше модель работает, тем больше она получает трафика
- Худшие модели убираем
- Плюсы:
  - Быстро переключаемся на модель, если она значительно лучше
- Минусы:
  - Сложнее, чем А/Б тесты

# Дополнительные материалы

- Automated Canary Analysis at Netflix with Kayenta
- Online Learning for Recommendations at Grubhub
- Overcoming Catastrophic Forgetting with Hard Attention to the Task
- Overcoming Catastrophic Forgetting with Unlabeled Data in the Wild

Все будет в телеграм-канале