

# Рекомендательные почтовые рассылки

---

# Задача

---

С вами общаются представители фанатского сайта с мультсериалами fox. Они хотят делать рассылки писем для своих пользователей, и для этого им нужна ваша помощь.

# Ограничения, не связанные с машинным обучением

---

1. Рассылки делаются только зарегистрированным пользователям
2. Нужно блюсти возрастные ограничения (6+, 12+, 16+, 18+ и т.д.). Возраст указывается при регистрации
3. Если человек смотрит какой-то сериал, не нужно предлагать следующие серии этого сериала

# Метрики качества

---

1. Конверсия письма в переход на слайд
2. Средняя возвращаемость (через сколько времени человек снова заходит на сайт и какая доля людей возвращается)
3. Среднее время на сайте и среднее число просмотренных минут серий
4. Время на сайте после перехода
5. Обычно такие сайты живут за счёт рекламы, можно смотреть на число кликов по рекламным блокам и число просмотров рекламных вставок

# Проверка качества

---

1. Статистически значимая оценка конверсии (доверительный интервал)
2. АБ тест для остальных метрик предыдущего слайда. Делим пользователей на две группы. Не забываем сделать АА тест, чтобы проверить, что разбиение нормальное. После чего запускаем АБ тест.

# Модель

---

Тут нет строгих требований к студентам, могут придумывать всё что угодно. Авторское решение такое:

По паре (пользователь, сериал/серия) предсказываем посмотрит ли он её в ближайшие  $y$  месяцев

Не забыть:

1. Придумываем для пользователя долгосрочные фичи, описывающие его интересы
2. По последним  $x$  месяцам строим фичи, описывающие краткосрочные интересы
3. Довольно много категориальных фичей, надо придумать как с этим жить (например, заменить на какие-нибудь счётчики)

# Калибровка модели

---

После первых экспериментов было бы неплохо построить новую модель, которая будет предсказывать конверсию письма в переход.

До этого мы придумали какую-то внутреннюю метрику качества (просмотр серии в будущем) и оптимизировали её. Но это не то, что нужно заказчику, ему нужны переходы на сайт, поэтому если есть возможно обучить модель на внешний критерий качества, нужно это сделать. В данном случае можно предсказывать будет переход или нет, или время которое будет проведено на сайте по переходу (0, если перехода не было).