

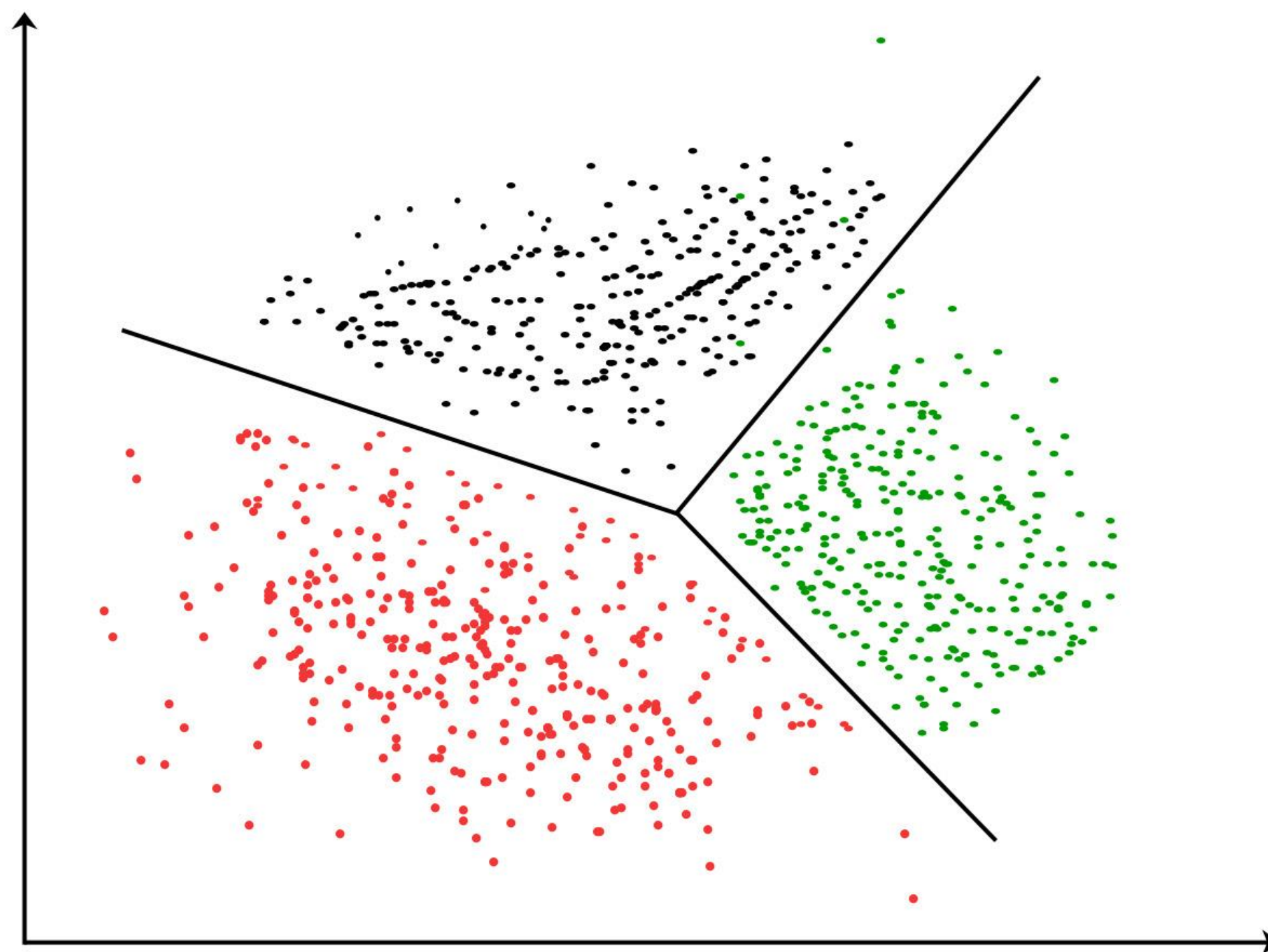


Сбор требований в Data Science-проектах

Евгений Виноградов, руководитель разработки хранилищ данных

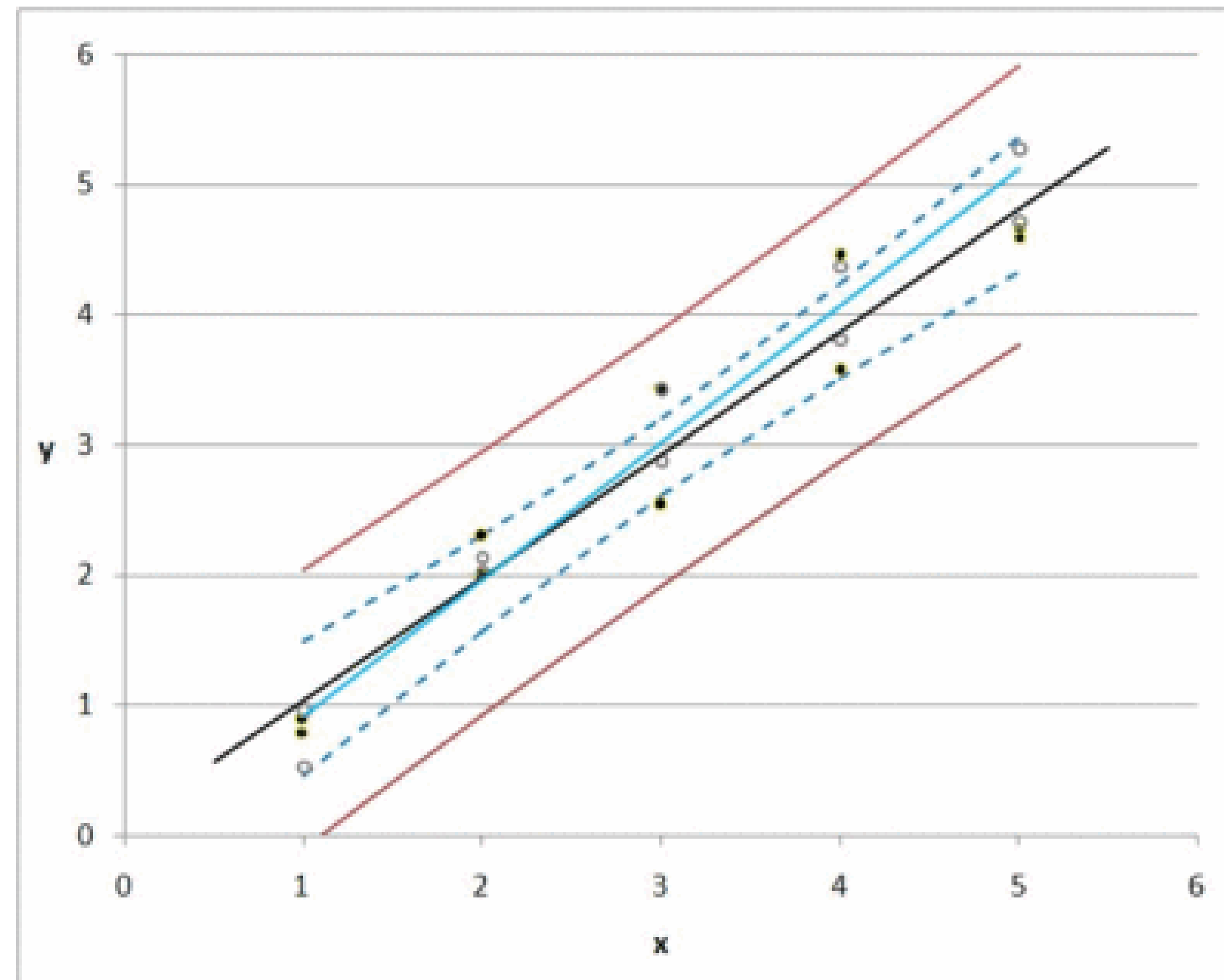
Модели

› Описательные



Модели

- › Описательные
- › Предсказательные



Модели

- › Описательные
- › Предсказательные
- › Агрегирующие



Модели

- › Описательные
- › Предсказательные
- › Агрегирующие
- › Классификаторы



Заказчики

› РО

› РМ

Заказчики

- › Маркетинг
- › Финансы
- › СБ

Сложности

Нет данных

Нет эксперта в предметной области

Не сформулирована проблема

Заказчик не понимает, что это RnD

Варианты взаимодействия

- › Есть проблема, и нужно подобрать вероятностное решение

Варианты взаимодействия

- › Есть проблема, и нужно подобрать детерминированное/гарантированное решение

Варианты взаимодействия

- › Заказчик озвучивает проблему и сам занимается ее решением, и ему нужна помощь в математике/технике

Варианты взаимодействия

- › Вы и есть заказчик

С чего начать?

- › С метрик успеха
- › Их нужно получить от заказчика
- › И перевести в DS-формат

Метрики

Некоторое количество придумали до вас

Метрики

$$\hat{Y} = 0$$

NEGATIVE

$$\hat{Y} = 1$$

POSITIVE

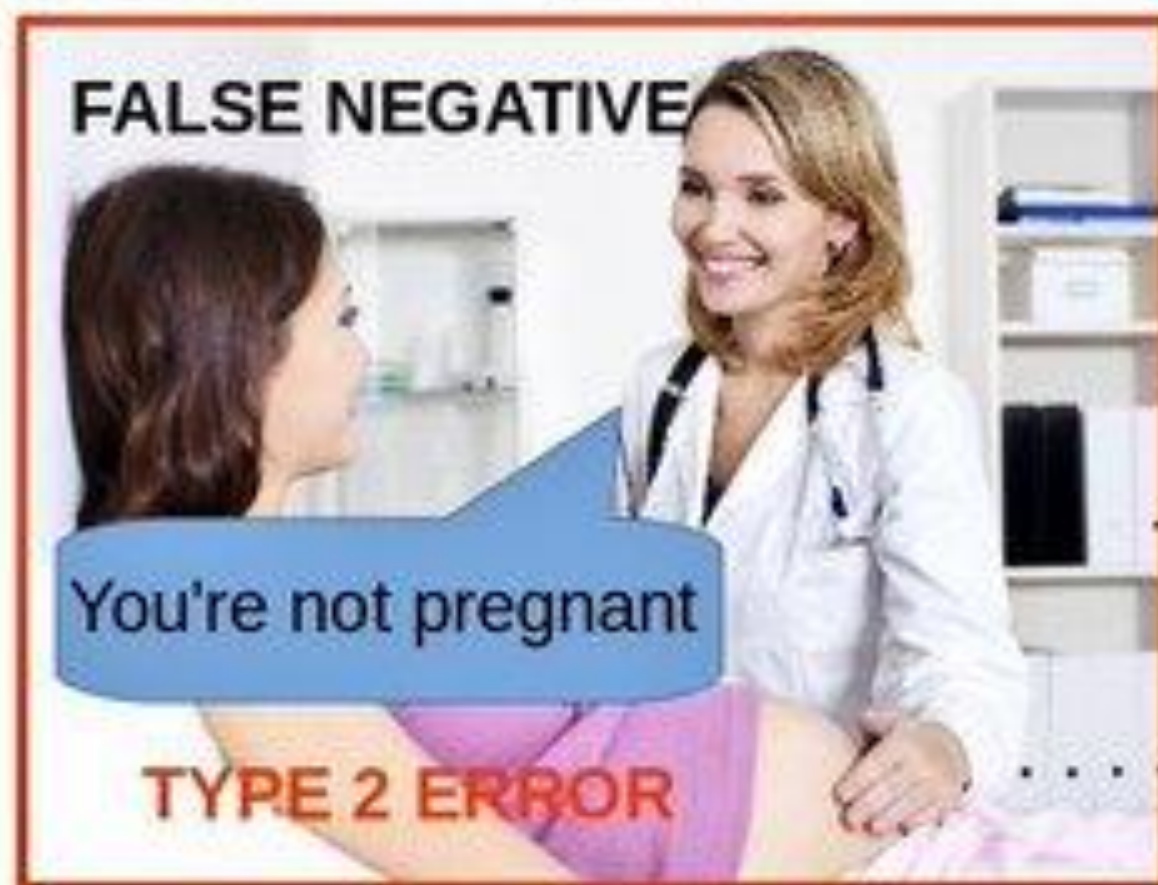
$$Y = 0$$

NOT PREGNANT



$$Y = 1$$

PREGNANT



Метрики

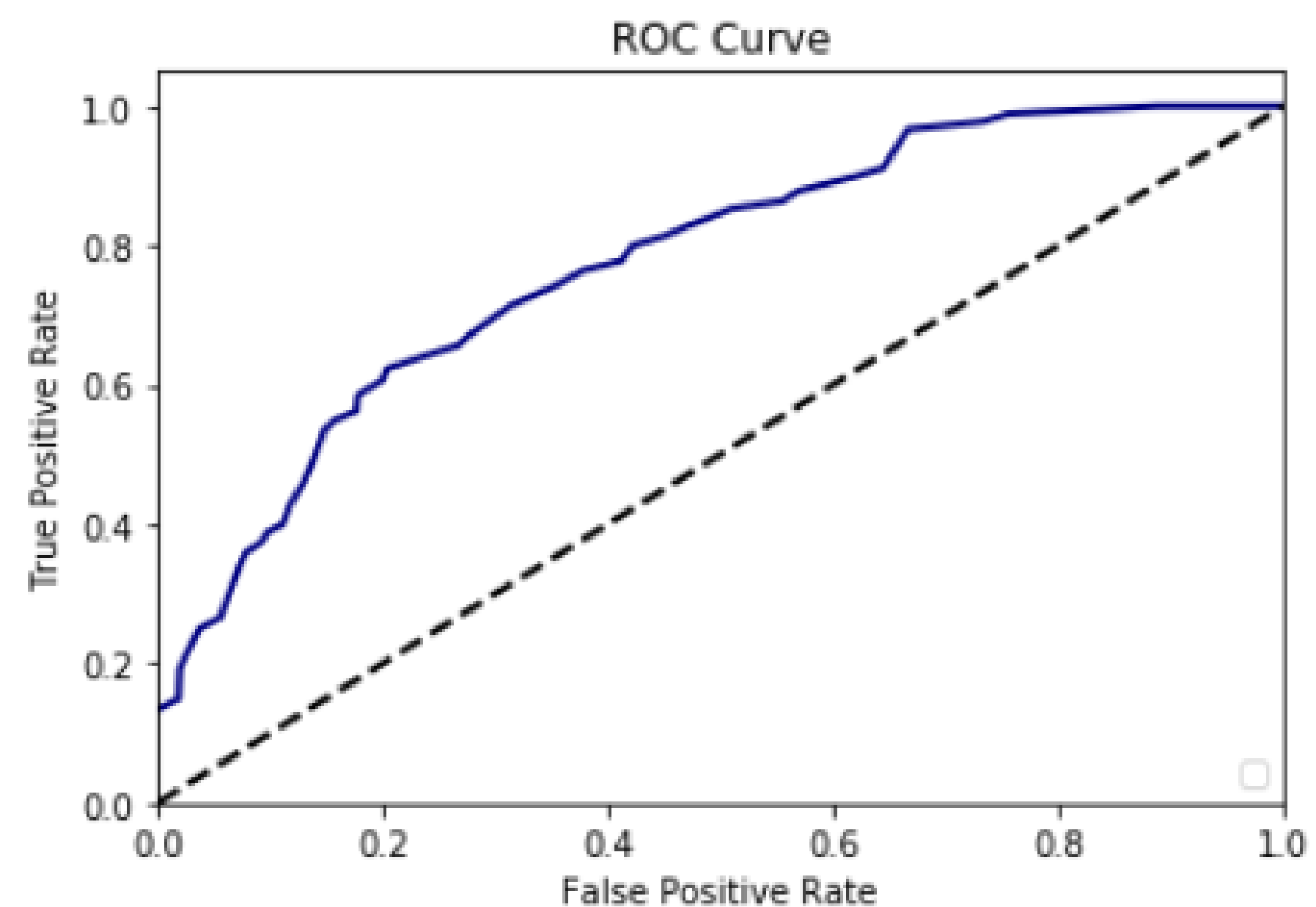
Некоторое количество придумали до вас

› Precision/recall

Метрики

Некоторое количество придумали до вас

- › Precision/recall
- › Гармоническое среднее
- › ROC-AUC



Метрики

Некоторое количество придумали до вас

- › Precision/recall
- › Гармоническое среднее
- › ROC-AUC

Что будем использовать?

Какие значения нас устроят?



А если не удалось достичь метрик?

- › Тюнинг
- › Смена модели

А если не удалось достичь метрик?

➤ ~~Тюнинг~~

➤ ~~Смена модели~~

› Обогащение данных

› Смена перспективы

› Смена метрик

Регулярное ревью

- › Точно ли нам все еще интересно именно решение исходной задачи?

Ключевые моменты

- › Если есть данные, гипотеза и специалист – может помочь Data Science
- › Метрики успеха крайне важны
- › Если не сработало – обогатите датасет!

Вопросы?

Контакты

Евгений Виноградов

Руководитель отдела разработки хранилищ
данных



jonny@yamoney.ru



+7(906)249-68-42