

Домашнее задание #1

Тема: Введение в задачу классификации. Постановка задачи и подготовка данных

Вопрос 1: Приведите по 2 примера, когда лучше максимизировать Precision, а когда Recall

Precision:

Precision важнее, чем Recall, когда стоит задача иметь меньше FP и больше FN.

1. Рекомендации Youtube, где ложно-отрицательные (FN) данные не опасны.
2. Во время апокалипсиса, необходимо принять как можно больше здоровых людей в свою безопасную зону, но по ошибке не пропустить зомби в безопасную зону. Так что если вдруг метод заставляет некоторых здоровых людей по ошибке не попасть в безопасную зону, то так тому и быть.

Recall:

1. Моделирование данных при диагностике рака. Все, что не учитывает ложно-отрицательные (FN) данные, является огромной ошибкой.
2. Определение террористов среди пассажиров (мы заинтересованы в поиске позитивных кейсов).

Вопрос 2: Почему мы используем F-меру, почему, например, нельзя просто взять среднее от Precision и Recall?

F1 сочетает в себе как Precision, так и Recall, так что эта метрика учитывает и одно, и другое. Причина, по которой F1 использует среднее гармоническое вместо усреднения обоих значений, заключается в том, что среднее гармоническое наказывает за экстремальные значения.

Пример:

- Precision = 1.0
- Recall = 0.0

Среднее значение будет 0.5, тем не менее, среднее гармоническое равно 0.0