

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

$$Precision = \frac{TP}{TP+FP}$$

лучше максимизировать при определении спама во входящей почте, что позволит избежать ложного попадания в папку спам важных писем (ложный спам > min);

или при принятии решения алгоритмами банка о выдаче кредита, что позволит уменьшить случаи невозвратов (ложный хороший заёмщик > min).

$$Recall = \frac{TP}{TP+FN}$$

лучше максимизировать при предсказании злокачественного новообразования по результатам анализов/МРТ, что позволит выявить максимум пациентов среди действительно больных (ложный не рак > min);

или при выявлении подозрительных операций алгоритмами банка, что позволит заблокировать большее количество операций среди всех действительно мошеннических (ложное не мошенничество > min).

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

$$F_{\beta} = (1 + \beta^2) \times \frac{precision \times recall}{(\beta^2 \times precision) + recall}$$

F-мера хороша тем, что позволяет одновременно учитывать Precision и Recall, но при этом ещё даёт возможность регулировать соотношение влияния этих метрик, через выбор β -коэффициента.