

# Все баллы умножаем на 4

## 3. Ловим мошенников (20 баллов)

Вы аналитик заявок на кредит в одном из крупных банков. Ваша задача — построить модель, предсказывающую, данная заявка мошенническая или нет. Данная модель позволит проводить проверки не каждый раз, как сейчас, а только когда она будет предсказывать мошенников. Вы знаете, что в среднем только 2% заявок мошеннические. Для каждого пункта данной задачи вы можете использовать информацию из предыдущих пунктов, но не из последующих.

(i). Допустим, вы построили модель, точность которой (accuracy, доля верных предсказаний) равна 95%. Может ли такая модель быть хорошей с точки зрения применимости к процессам проверки? Может ли такая модель быть плохой с точки зрения применимости к процессам проверки?

Решение:

Давайте рассмотрим все наблюдения, в которых модель сделала неверное предсказание. Допустим, там только немошенники – это значит, что если мы будем использовать модель, то ни один мошенник не пройдет, а количество проверок уменьшится. То есть, модель хорошая, так как позволяет проверять меньше заявок (лишь 5%) без потери качества (не выдавая кредиты мошенникам).

Рассмотрим противоположный случай – когда модель плохо предсказывает всех мошенников. Заметим, что у тривиальной «модели», которая считает все заявки не мошенническими, точность будет 98%. Тогда полученная модель будет хуже тривиальной и если ее применять, то все мошенники смогут получить кредит, а проводимые проверки будут бесполезны, так как проверяют только хороших заявителей.

Критерии:

Мах – **32 балла**

Обоснование с примером для хорошей модели – **16 баллов**

Обоснование для плохой модели – **16 баллов**, если используется доля мошенников. Без использования – **0 или 8 баллов**

(ii). Ниже вы можете увидеть результаты предсказания модели на выборке заявок за прошедший месяц.

		Результат проверки	
		Мошенник	Не мошенник
Предсказание модели	Мошенник	77	374
	Не мошенник	127	9487

Посчитайте точность (accuracy, доля верных предсказаний) вашей модели. Посчитайте долю мошенников в выборке.

Решение:

Точность = (верно предсказанные)/(все заявки) =  $(77+9487)/(77+374+127+9487) = 0.9502$

Доля мошенников = (мошенники)/(все заявки) =  $(77+127)/(77+374+127+9487) = 0.0203$

Критерии:

Мак – **4 балла**

По **2 балла** за ответ

(iii). Ваш коллега посчитал, что каждая проверка обходится нам в 100 ₽, а из-за одного мошенника мы теряем 500 ₽.

Сколько мы заработаем (сэкономим), если будем использовать модель?

Сколько мы заработаем (сэкономим), если вместо внедрения модели перестанем проводить проверки вообще?

Решение:

Предположим, что поток заявок за предыдущий месяц репрезентативен

Тогда при использовании модели мы будем проводить на 127+9487 проверок меньше, при этом пройдет 127 мошенников. Суммарный экономический эффект =  $(127+9487)*100 - 127*500 = 897900$  рублей экономии в месяц

При прекращении проверок мы будем проводить на 77+374+127+9487 проверок меньше, при этом пройдет 77+127 мошенников. Суммарный экономический эффект =  $(77+374+127+9487)*100 - (77+127)*500 = 904500$  рублей экономии в месяц

Критерии:

Мак – **16 баллов**

Обосновано использование предыдущего месяца – **8 баллов**

Посчитано для модели – **4 балла**

Посчитано для отключения проверок – **4 балла**

(iv). Что будет, если действительно отключить все проверки начиная со следующего месяца? Значит ли это, что заработаем столько, сколько было посчитано в предыдущем пункте? Можем ли мы повлиять на поток заявок? Можем ли мы в итоге потерять деньги, а не заработать?

Решение:

Если отключить все проверки, то мошенники, которые подавали заявку на получение кредита и раньше отлавливались ручными проверками, теперь смогут получить кредит. Значит, все больше и чаще мошенники будут пользоваться данной лазейкой, доля мошенников в потоке вырастет, вычисления предыдущего пункта перестанут быть верными для новых месяцев, и мы в итоге можем потерять деньги от нововведения, если доля мошенников станет слишком большой.

Критерии:

Мах – **28 баллов**

Описано влияние нашего решения на поток заявок с точки зрения мошенников – **16 баллов**

Сделан вывод об изменении потока – **4 балла**

Обоснована возможность потери денег и описаны причины столь резкого изменения потока заявок – **8 баллов**