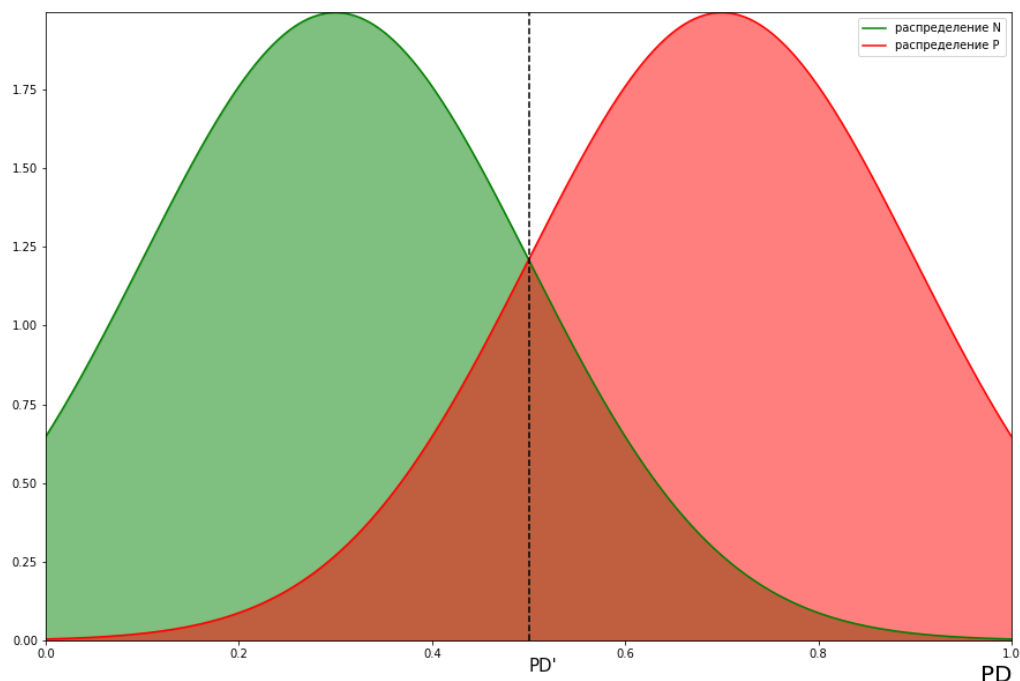


Соответственно доля ошибок и правильных предсказаний модели при предсказании появления исследуемого признака.

Как правило, модель не позволяет сразу однозначно определить класс, потому, при варьировании точки отсечения  $PD'$  будут получены различные показатели  $FPR$  и  $TPR$ . Для лучшего понимания взгляните на рисунок 2.1.



**Рисунок 2.1 – Распределение реальных значений P и N**

Примечание – Источник: собственная разработка.

На оси абсцисс отложены вероятности наличия признака предсказываемые моделью, зеленой площадью представлено предполагаемое распределение клиентов с отсутствием исследуемого признака, красной – с присудившем. Предсказанные вероятности до  $PD'$  относят соответствующие наблюдения к категории без проявления исследуемого признака, после  $PD'$  – к категории с проявлением признака.

Изменяя значение  $PD'$  можно влиять на соотношения в таблице 3.1. Увеличение  $PD'$  будет приводить к тому, что для большего числа наблюдений будет предсказано отсутствующий исследуемый признак (рост  $N^*$  и спад  $P^*$ ). В случае снижения  $PD'$  все будет работать наоборот.

Визуализация идей изложенных выше представлена на рисунках Е.1 и Е.2, они представляют строгую и лояльную модели соответственно. Глядя на эти рисунки сделаем некоторые выводы.

Обычно при моделировании используются показатели  $TP$ ,  $FP$  и их относительные аналоги  $TPR$ ,  $FPR$ , потому, сконцентрируемся именно на них.

Начнем из ситуации  $PD' = 1$  – совершенно лояльной модели: все наблюдения с проявившимся признаком будут включены в  $FP$ , все наблюдения без признака будут включены в  $TN$ . Из того следует, что  $FPR = 0$ ,  $TPR = 0$ .