

Издержки ошибок классификации (Classification cost error)

Синонимы: Стоимость ошибок классификации

Loginom: Логистическая регрессия (обработчик), Нейросеть (классификация).
(обработчик).

Издержки (потери, убытки) от ошибок классификации, допущенных аналитической моделью. Такие ошибки могут приводить к неверному принятию решений в бизнесе, а это, в свою очередь, влечет к материальным и финансовым издержкам, которые в аналитических технологиях Data Mining часто называют издержками ошибки классификации.

Например, целью работы модели для оценки кредитоспособности служит выявление добросовестных и недобросовестных клиентов. Данная задача известна как бинарная классификация. Закономерен вопрос: что лучше — принять добросовестного клиента за недобросовестного (ложно-отрицательная ошибка) или наоборот (ложно-положительная ошибка)?

В первом случае мы теряем только проценты по кредиту, который не был выдан, а во втором — всю сумму, которую получил недобросовестный заемщик, если он не сможет вернуть кредит.

Иными словами, издержки ошибок второго вида больше. Это значит, что при построении модели мы должны минимизировать вероятность появления ошибок, которые приводят к наибольшим потерям.

Типичными примерами, когда издержки неодинаковы для разных типов ошибок, являются:

- кредитный скоринг — издержки выдачи кредита недобросовестным заемщикам существенно выше, чем потери бизнеса из-за отказа в выдаче кредита добросовестным клиентам;
- поиск очагов нефтяного загрязнения — издержки пропуска реального нефтяного пятна существенно выше, чем ложной тревоги;
- техническая диагностика — издержки неправильной идентификации проблемы существенно меньше, чем возможные потери от ее пропуска;
- директ-маркетинг (рассылка прямой почтовой рекламы) — затраты на отправку «макулатурной почты» клиентам, которые не отвечают на нее, существенно меньше, чем потери бизнеса из-за пропуска потенциальных клиентов.

Вообще, в экономике и бизнесе трудно найти приложение, в котором издержки различных типов ошибок были бы одинаковыми, поэтому использование моделей, учитывающих издержки классификации, очень актуально.