

Розничный кредитный риск

Наталья Васильева

Высшая Школа Экономики
nvasileva.1609@gmail.com

Финансовые Технологии и Анализ Данных
2 февраля 2024 г.

- 1 Розничный кредитный риск
- 2 Кредитный скоринг
- 3 Особенности задачи кредитного скоринга
 - Пропуски в данных
 - Выбор данных для моделирования

Розничный риск

основная особенность – массовость; каждый отдельный кредит пренебрежимо мал (<100 тыс. \$)

- агрегация, использование статистики и предельных теорем
- диверсификация

Примеры:

- кредитные карты;
- автокредиты, кредиты на образование;
- потребительские кредиты;
- ипотечные кредиты;
- кредиты малому бизнесу.

1 Кредитный скоринг – кредитный анализ

- принять решение по заявке на кредит;
- отсеять слишком рискованные заявки;
- модифицировать условия кредита;
- не угадывать, кто вернёт кредит.

2 Оценка риска и аллокация капитала

- по имеющемуся портфелю кредитов оценить риск и заложить капитал на его покрытие

3 что дальше делать с кредитом?

Проблемы розничного кредитного риска

- не по всем кредитным продуктам достаточно исторических данных;
- смена экономической обстановки может сильно повлиять на исходы;
- смена политической, юридической и социальной обстановки также может влиять (социальная примелемость дефолта);
- все проблемы приобретают масштабный характер из-за пакетной обработки;
- проблемы видны с заметным лагом.

Кредитный риск – не единственный риск в розничном кредитовании

- процентный риск (управляется казначейством);
- риск переоценки активов, как кредитов, так и залогов (управляется казначейством);
- бизнес-риск;
- репутационный риск.

Кредитный скоринг

Классическая задача ML

На входе: анкета заёмщика и дополнительная информация (результаты проверок)

На выходе: решение по кредиту.

Качественное отличие:

- участие человека в принятии решения;
- необходимость покупки доп. информации для принятия решения;
- вспомогательная модель для оценки необходимости покупки доп. информации.

Упрощённая схема скорингового процесса

Классическая задача ML

На входе: заявка

Обработка:

- прескоринг (политика банка, полнота документа);
- автоматизированная скоринговая система – оценка риска;
- СБ (вспомогательная модель);
- кредитный отдел (экспертная оценка).

Принятие решения:

- 1 отказ;
- 2 одобрение;
- 3 предложение кастомизации.

Оценка риска

Скоринговые модели

Скоринговый балл = $f(\text{анкета, доп.инфо})$

- высокий балл – принять;
- низкий балл – отказать;
- серая зона – принятие решения экспертом.

Дополнительная информация может закупаться всегда или только в случае необходимости.

- вспомогательная модель: принять, отказать, закупить доп. информацию.

Типичная задача ML.

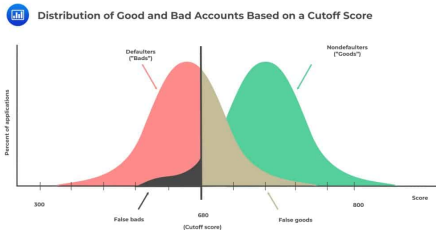
Виды дополнительной информации

- если клиент банка – анализ истории взаимодействия и истории транзакций (зарплатные клиенты);
- активность пользователя в экосистеме банка;
- данные БКИ;
- данные от сотовых компаний и социальных сетей;
- готовый скоринг вместо исходной информации (generic models).

Виды дополнительной информации

- если клиент банка – анализ истории взаимодействия и истории транзакций (зарплатные клиенты);
- активность пользователя в экосистеме банка;
- данные БКИ;
- данные от сотовых компаний и социальных сетей;
- готовый скоринг вместо исходной информации (generic models).

Принятие решения на основе скорингового балла



- высокий балл – принять;
- низкий балл – отказать;
- серая зона – принятие решения экспертом.

Качество моделей

Виды ошибок

| | | Observed (Y_i) | |
|--------------------------|--------|--|--|
| | | Default | No-Default |
| Decision (\hat{Y}_i) | Reject | True Positive | FP (I type) |
| | Accept | FN (II type) | True Negative |
| | | Positive (P) Total # of default (bad) clients | Negative (N) Total # of creditworthy (good) clients |

positive – дефолт.

❶ ошибка I рода – FP;

- отказ в кредите "хорошему" заёмщику;
- нет статистики (сквозное одобрение).

❷ ошибка II рода – FN;

- выдача кредита "плохому" заёмщику;
- статистика есть.

не только аппликативный (заявочный):

- 1 behavioral scoring – вероятность дефолта в течение жизни кредита;
- 2 fraud scoring – выявление мошеннических операций;
- 3 collection scoring – меры работы с должником применять при просрочке (реструктуризация);
- 4 revenue scoring – прибыльность клиента;
- 5 response scoring – целесообразность специального предложения клиенту;
- 6 attrition scoring – отток клиентов.

Особенности задачи кредитного скоринга

- ① разные целевые переменные при выдаче кредита;
 - горизонт вызревания
 - досрочное погашение
 - прибыльность
- ② достоверность анкетных данных (вспомогательная модель);
- ③ пропуски в данных;
- ④ выбор переменных для анализа;
- ⑤ выбор формы модели.

Пропуски в данных

Классификация пропусков

Анкетные данные часто бывают пропущенными. Важен механизм их появления.

- факт пропуска может нести информацию

Классификация пропусков

- 1 Missing Completely At Random (MCAR);
- 2 Missing At Random (MAR);
- 3 Missing Not At Random (MNAR);

Классификация пропусков

Missing Completely At Random (MCAR)

Факт пропуска в принципе не зависит от содержания данных. Записи с полностью доступными данными – случайная и независимая подвыборка:

- если немного, можно удалить данные с пропусками;
- если много, можно удалить весь признак.

Классификация пропусков

Missing At Random (MAR)

Факт пропуска может зависеть от содержания записи, но только от наблюдаемых значений, а не от пропущенных.

Пропуски вида "Not Applicable"

- информация о средней зарплате за последние 6 месяцев будет пропущена, если последние 6 месяцев респодент нигде не работал;
- вероятность пропуска зависит от наблюдаемых параметров анкеты, пропуск коррелирован с другим полем;
- исключение таких записей может привести к искажению выборки;
- необходим подбор адекватной модели и метода оценки параметров.

Классификация пропусков

Missing Not At Random (MNAR)

Факт пропуска зависит от пропущенных данных.

- намеренное утаивание информации в анкете;
- отклонение заявки СБ;
- цензурирование: значение отсутствует из-за масштаба (досрочное погашение – цензурированная целевая переменная).

Не следует исключать!

Классификация пропусков

Нестационарные модели

Вероятность пропусков меняется со временем (с календарным временем и с возрастом кредита – МОВ).

- сильный эффект в ипотечных кредитах;
- ранее поле было не обязательным, стало обязательным;
- более качественная отчётность;
- используют внешние данные (макроэкономические показатели);
- без внешней информации уход от нестационарности: квантили; вспомогательные признаки – кластеры.

Не следует исключать!

Что делать с пропусками?

- исключать
 - если данные не MCAR, это обычно приводит к ошибкам в оценках
- использовать модели (и численные методы), учитывающие пропуски
 - например, Buckley-James Regression, цензурированные данные
- рассмотреть пропуск как ещё один вариант заполнения поля
 - не все модели с этим работают
- заполнять
 - конкретным значением: EM-алгоритм
 - условное мат. ожидание (при условии других наблюдаемых полей – вспомогательная модель
 - выборкой из условного распределения (multiple imputation, байесовские методы)
- дополнительные признаки на основе пропусков.

Заявка подана, анкета заполнена, но отказ по кредиту.

- ошибка I рода
- характеристика входного потока заявок
- потеря информации при построении модели только по выданным кредитам

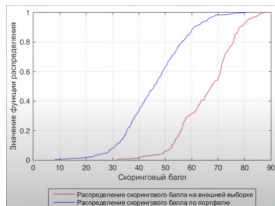
Выбор данных для анализа

Важно при сборке выборки понимать механизмы фильтрации и отбора.

- кредиты сотрудникам
- рекламные акции (предварительное одобрение)
- невозвраты по финансовым и по нефинансовым причинам (мошенничество, смерть)
- досрочное погашение кредита

Смысл данных и их релевантность может меняться на горизонте выборки (деноминация).

Качество данных (СБ, обязательные для заполнения).



Определение "дефолта"

- просрочка платежа DPD на N дней (часто 30/90/180) на горизонте 1 года
- банкротство, ликвидация
- реструктуризация

Если дефолтов мало:

- поменять определение дефолта
- выбор модели, сильнее регуляризовать
- выбор прокси (рейтинг – используется в корпоративных рисках)