

Разбор решения

BNP Paribas Cardif Claims Management

Николаев Владимир Владимирович

ВМК МГУ 317 группа

22 апреля 2015 г.

- Есть какие-то анонимизированные данные.
- Нужно определить 0/1 в ответе.
- Классы немного не сбалансированы: среднее 0.76.
- Метрика качества: **LogLoss**.

- Сумма пропущенных значений в строке
- Сумма нулей в строке

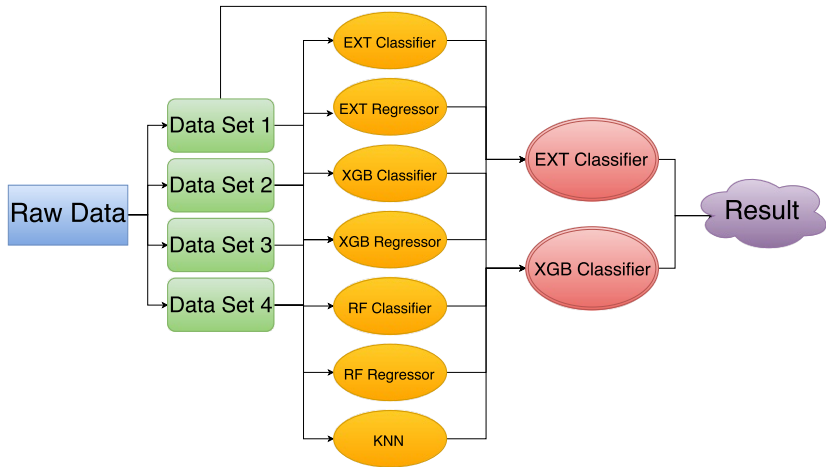


Рис. 1 : Решение

- За 2 недели до конца соревнования в скриптах есть много полезной информации
- Двухуровневые модели хорошо работают
- Полезно сохранять модели (или хотя бы их предсказания)
- Усреднение xgboost'а эквивалентно калибровке под вероятности