

Предсказание аварий

Тренировка для начинающих

Задача

- Вероятность аварии по данным анкеты водителя
- <https://goo.gl/31jw5d>
- Small data
- Случайное разбиение train/test
- ROC AUC метрика
- Числовые и категориальные признаки

Простая схема

- Загрузить данные
- “Почистить”
- “Подготовить” признаки для модели
- Обучить модель на обучающей части
- Применить модель на тестовой части
- Собственная оценка качества решения
- Результат отправить в систему контеста

Решение А

- Паблик : ~ 0.842
- Идеи:
 - ♦ Использовать числовые признаки
 - ♦ Использовать категориальные признаки
 - ♦ RF на большое число деревьев

Решение В

- Паблик : ~ 0.850 ; 5-fold CV ~ 0.823
- Идеи:
 - ◆ Использовать CV
 - ◆ Настроить RF и XGBoost
 - ◆ Блендинг моделей

Вопросы?

Можно приступать!