

Подбор признаков

Домашнее задание

Необходимо как можно выше подняться в лидерборде соревнования Titanic.

В задании необходимо выполнить следующие пункты:

- Описать все имеющиеся признаки словами. По каждому признаку составить список особенностей. При необходимости построить соответствующие визуализации.
- Сгенерировать идеи по преобразованию этих признаков
- Составить план по преобразованию признаков - что, как и в каком порядке кодировать
- Составить список моделей для перебора. Начать с простых (константное преобразование, kpp). Перейти к более сложным - логистическая регрессия. Описание составить в виде "модель, какие параметры нужно подбирать".
Опционально. Попробовать наивный байес, SVM, дерево решений и случайный лес.
- Описать процесс кросс-валидации и подбора параметров каждой модели из плана (сделать акцент на кросс-валидации, нельзя подбирать параметры ориентируясь на лидерборд)
- Реализовать план. Лучшую модель засабмитить.
- Повторять, пока есть желание
- **Опционально.** Реализовать все через pipeline.

Критерии оценки.

- Проработка признаков, работа с числовыми, категориальными, текстовыми признаками
- План по применению моделей
- Все модели подобраны через кросс-валидацию
- Есть сабмит не хуже сабмита занятия (прислать свой ник, и ссылку, если есть такая возможность)