

Сбор данных

Кросс-валидация

```
In from sklearn.model_selection import cross_val_score  
cross_val_score(model, features, target, cv=3)  
  
# model – необученная модель для кросс-валидации;  
# features – признаки;  
# target – целевой признак;  
# cv – число блоков для кросс-валидации (по умолчанию их три).
```

Словарь

«Кроулер» или «Скрейпер»

специальные программы, предназначенные для сбора данных с сайтов

Неразмеченные данные

данные, в которых отсутствуют значения целевого признака

Разметка, аннотирование данных

процесс определения значений целевого признака

Размеченные данные

данные, в которых есть значения целевого признака

Утечка целевого признака

ситуация, когда в признаки случайно попадает информация о целевом признаке

Кросс-валидация

способ обучения и тестирования модели, в котором обучающая выборка делится на K частей. На каждом из K этапов i -ый блок становится валидационной выборкой, а остальные - тренировочной