

# Категоризация данных (Data categorization)

Разделы: [Бизнес-задачи](#)

В [искусственном интеллекте](#) и [анализе данных](#) категоризацией называют задачу отнесения [наблюдений](#) к одной из групп, называемых категориями. В целом задача категоризации похожа на [задачу классификации](#), но имеются следующие отличия:

- категория определяет лишь некоторые общие свойства объектов;
- задача категоризации менее определена, чем задача классификации;
- границы категорий, в отличие от [классов](#), являются нечеткими, а сама категория определяется не формально, а сравнением с другими категориями.

Задача категоризации является более сложной, чем классификации, поскольку классы предварительно заданы, а категории приходится формировать. В случае [обучения с учителем](#) формирование категорий происходит методом проб и ошибок на основе примеров с известными ответами, предоставляемыми экспертом.

Задачей системы-категоризатора является формирование обобщающих [признаков](#) на основе [обучающих примеров](#). При увеличении числа примеров несущественные, случайные значения признаков сглаживаются, а часто встречающиеся — усиливаются, при этом происходит постепенное уточнение границ категорий.