

# Несбалансированные выборки

**Андрей Мещеряков**  
Senior Data Scientist в EPAM Systems

Skillbox

Несбалансированные выборки

# Андрей Мещеряков

## 5+ лет опыта

в сфере Data Science  
и Machine Learning

## Создавал

решения на основе машинного  
обучения для ритейла  
и маркетинга



# Цель модуля

Узнать о подходах к работе  
с несбалансированными данными.

# Upsampling. Downsampling

# Цель

Разобрать методы выравнивания  
данных upsampling и downsampling.

# Выводы

Узнали принцип стратификации при разделении датасета, благодаря которому степень дисбаланса классов после разделения на выборки сохраняется.

Также разобрали 2 метода балансировки данных, upsampling и downsampling, которые позволяют выровнять количество записей позитивного и негативного классов.