

Основные понятия модуля

EDA — разведывательный анализ данных — анализ основных свойств данных, нахождение в них закономерностей, характера и свойств данных.

EDA включает в себя:

- Проектирование признаков
- Отбор признаков
- Кодирование признаков
- Построение и проверка статистических гипотез

Цикл машинного обучения

Кнопка	Действие
Постановка проблемы	Жизненный цикл нашей модели начинается с определения проблемы бизнеса
Сбор данных	На этапе сбора необходимо собрать как можно больше любых данных
Очистка данных	На этапе очистки данных мы определяем пропущенные значения, аномалии и выбросы в данных. Подготавливаем набор данных к анализу
EDA	Исследование признаков, извлечение полезной информации
Моделирование	Построение предсказательных моделей, проверка качества модели
Продакшен	Вывод модели в продакшен

Направления профессии

Профессия	Роль
Дата-сайентист	Создает рекомендательные алгоритмы и прогнозные модели
Дата-аналитик	Проводит предварительный анализ данных и их очистку, ищет закономерности в данных и ответы на запросы бизнеса
Дата-инженер	Обеспечивает сбор и хранение данных. Занимается внедрением модели

Инструменты автоматической визуализации

Библиотека	Установка
Pandas-profiling	<code>pip install pandas-profiling</code>
Sweetviz	<code>pip install sweetviz</code>
D-tale	<code>pip install dtale</code>