**Вопросы:**

1 Какие форматы данных используются при обучении моделей?

2 Посмотреть и повторить за спикером практическую часть.

**Ответ:**

1. Базовые, самые распространённые форматы- CSV/TSV. Довольно простой формат, преобразует входящие данные в таблицы.   
   JSON - это уже иерархический формат. В него можно довольно легко вносить дополнения, делать подпункты.

Message Pack - логический наследник JSON. Позволяет объединять данные более компактно, при этом сохраняет все достоинства последнего.

Feather - ещё один наследник JSON, позволяющий добавлять метаданные прямо в структуру. Способен работать с пакетами.

HDF5 - очень удобный формат, иерархический, позволяющий делать сколь угодно большие ветвления. Заточен на быстрые выборки из большого объёма данных.

Parquet - самый продвинутый на сегодняшний момент формат. Хранит данные поколоночно, что позволяет довольно легко изменить любые данные в любом месте.

1. <https://colab.research.google.com/drive/1MFAmEOBWwUJ95zbWGItuSRpwJfcYJXPE?usp=sharing>