



Введение в ML

Студент Сербин Г.Э.

Группа ML-12

Введение в анализ данных и машинное обучение

✓ Теория:

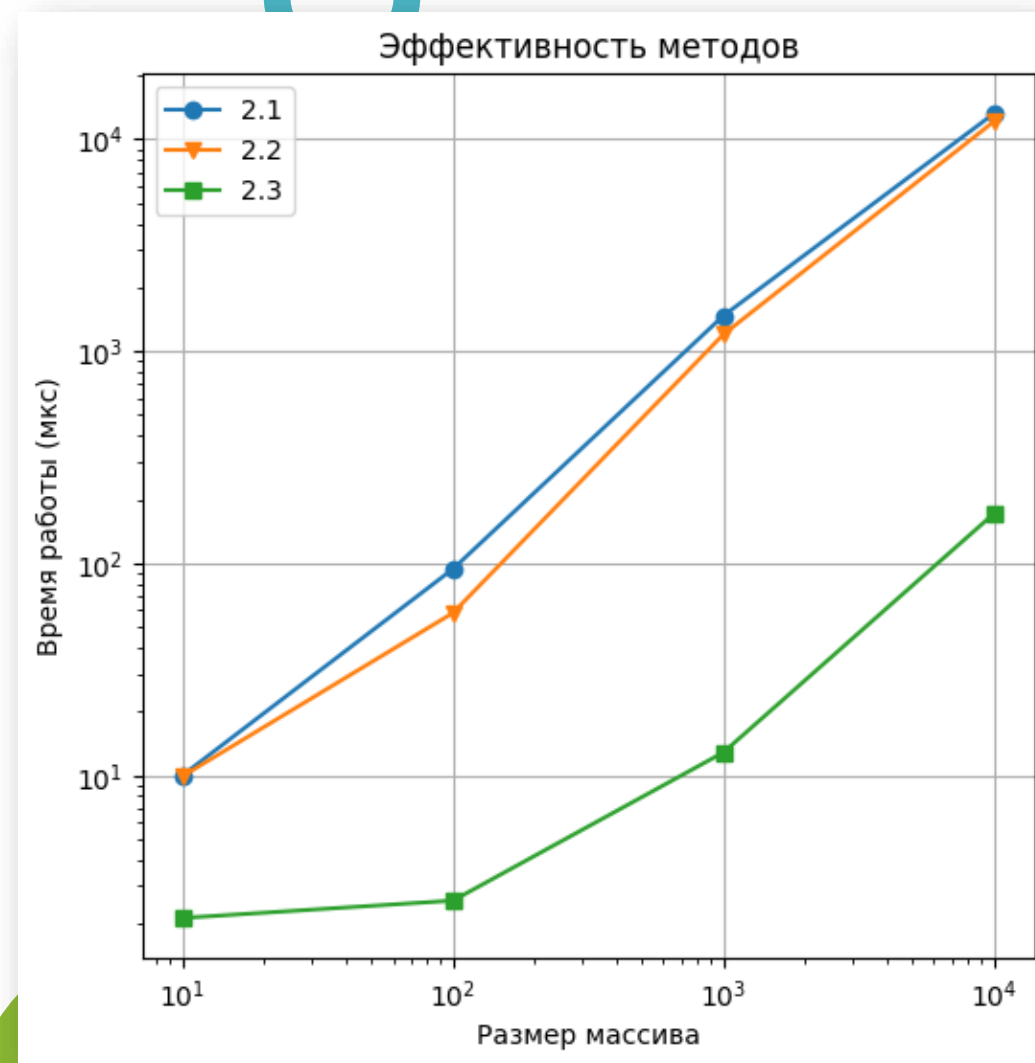
Что такое ML и какие задачи в нем решаются

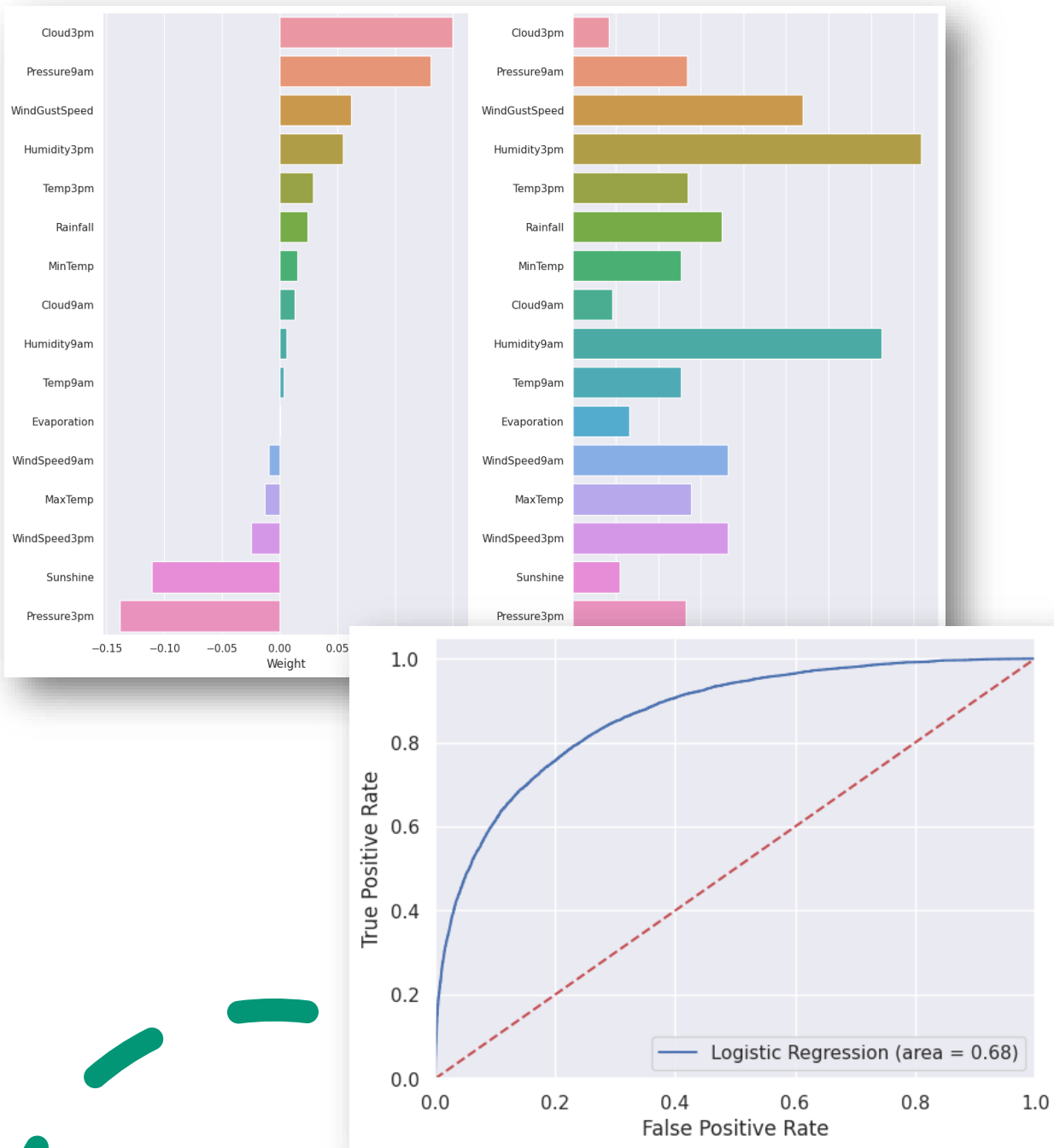
✓ Практика:

Библиотеки Python (matplotlib, pandas, numpy)

✓ Выводы:

При больших N код, использующий циклы, работает очень долго.





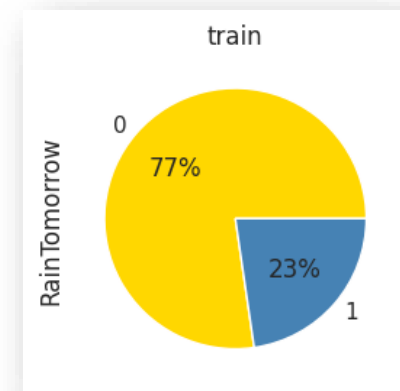
Задачи классификации и регрессии

✓ Теория:

Предобработка данных, линейные модели, kNN, оценка качества работы алгоритмов

✓ Практика:

Предсказание погоды



✓ Выводы:

Предобработка существенно влияет на качество модели

Работа с текстовыми данными I

✓ Теория:

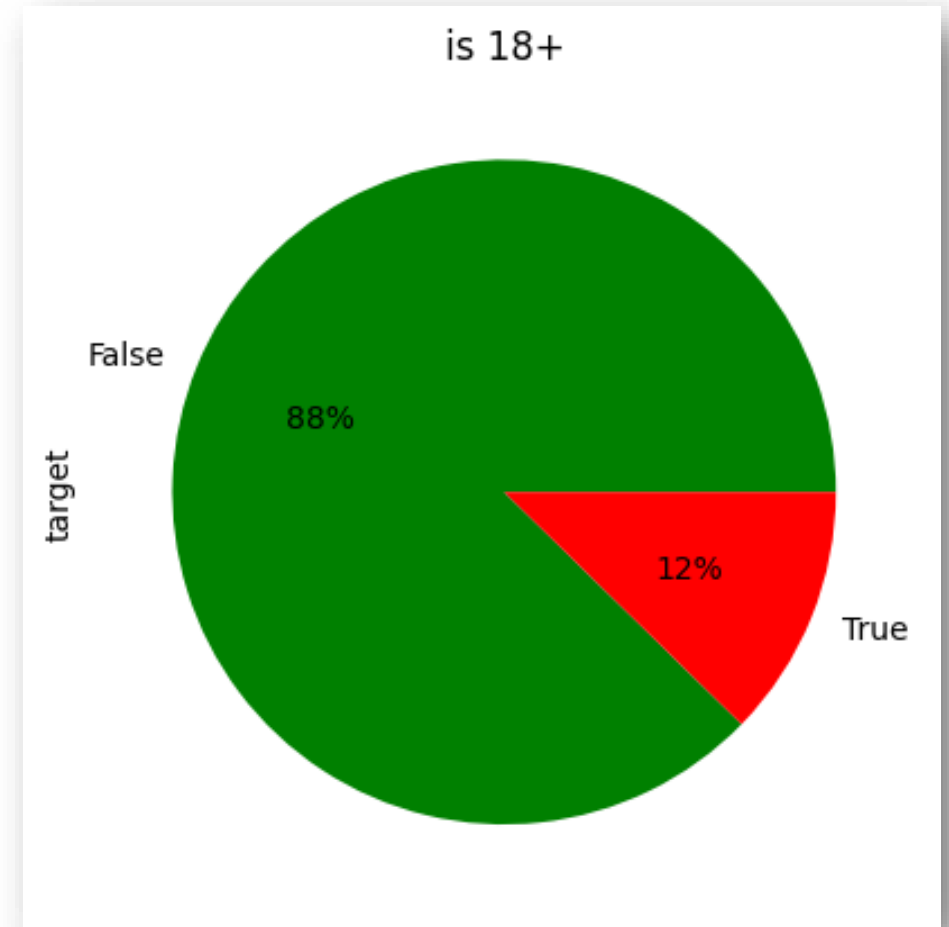
Bag of words, токенизация, стоп-слова, нормализация, стемминг, tf-idf, косинусная близость, embeddings

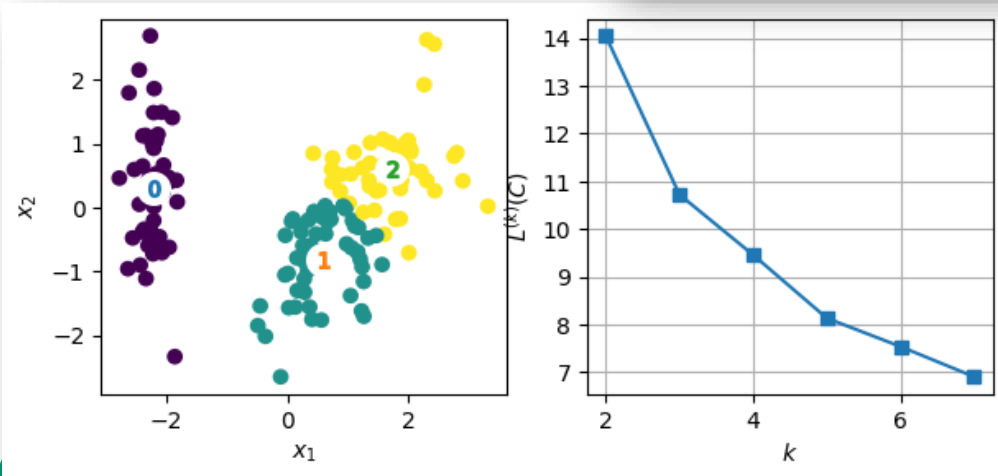
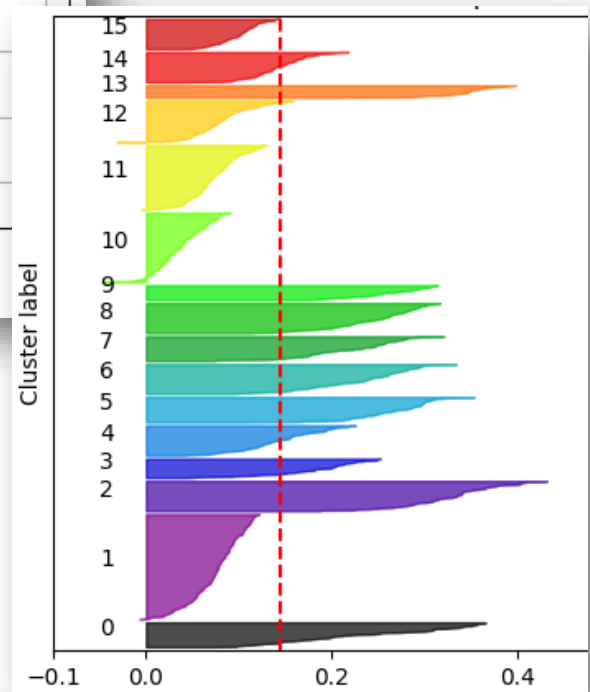
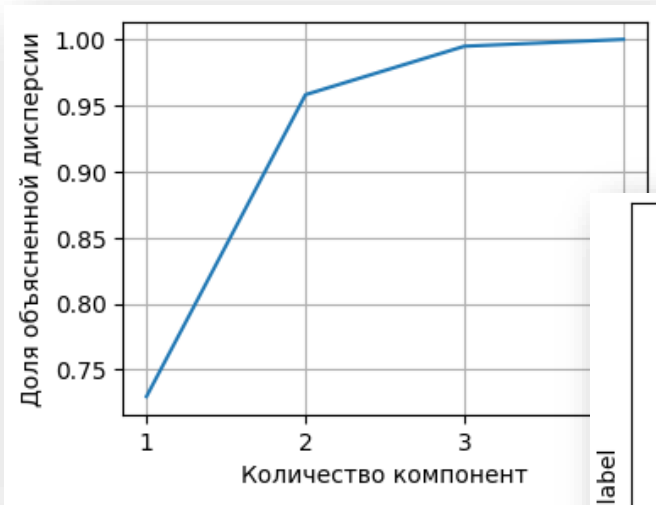
✓ Практика:

Классификация сайтов

✓ Выводы:

Важна предобработка текста





Обучение без учителя

✓ Теория:

Понижение размерности: PCA, t-SNE

Кластеризация: k-means, DBSCAN

✓ Практика:

Задача про датчики на руках, кластеризация текстов

✓ Выводы:

Можно делать красивые визуализации и иногда улучшать качество модели

Ансамбли моделей

✓ Теория:

Деревья решений: information gain, критерии останова, работа с категориальными признаками

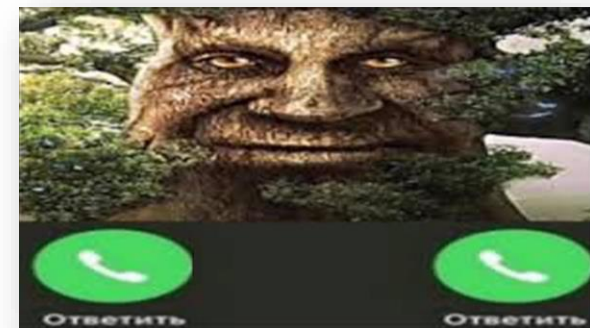
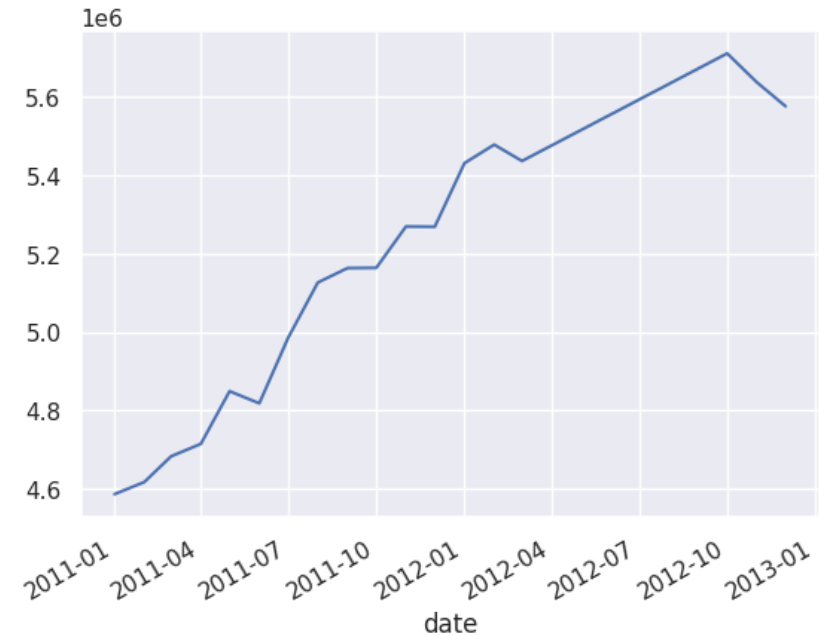
Ансамбли: случайный лес, градиентный бустинг

✓ Практика:

Прогнозирование стоимости жилья

✓ Выводы:

Деревья легко переобучаются и плохо экстраполируют



Работа с текстовыми данными II

✓ Теория:

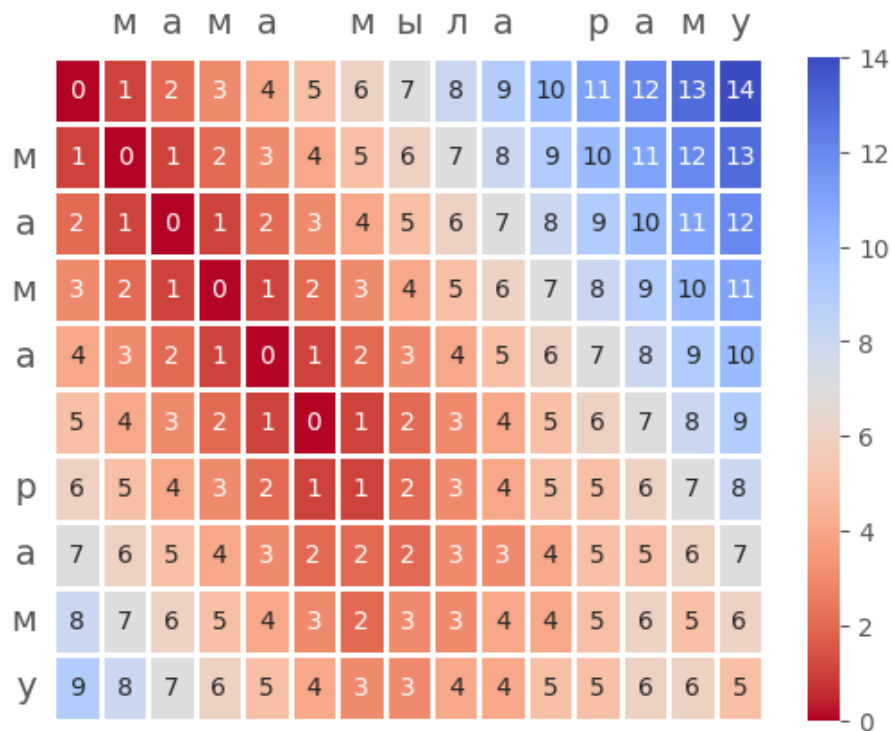
Self-supervised learning,
Word2Vec, Language Models,
Probability chain rule

✓ Практика:

Noisy text aggregation,
Levenstein-distance, rover,
error-correction, rescoring

✓ Выводы:

Неразмеченные данные могут быть очень
полезны



Рекомендательные системы

✓ Теория:

Explicit/implicit ratings

Кластеризация пользователей

Модель совстречаемости

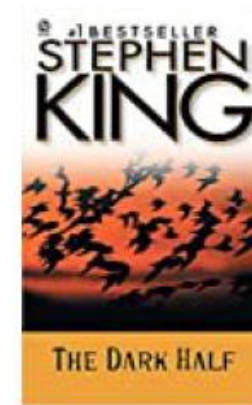
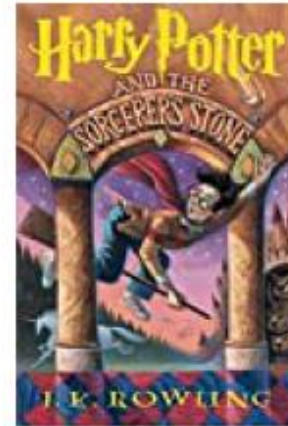
Коллаборативная фильтрация

✓ Практика:

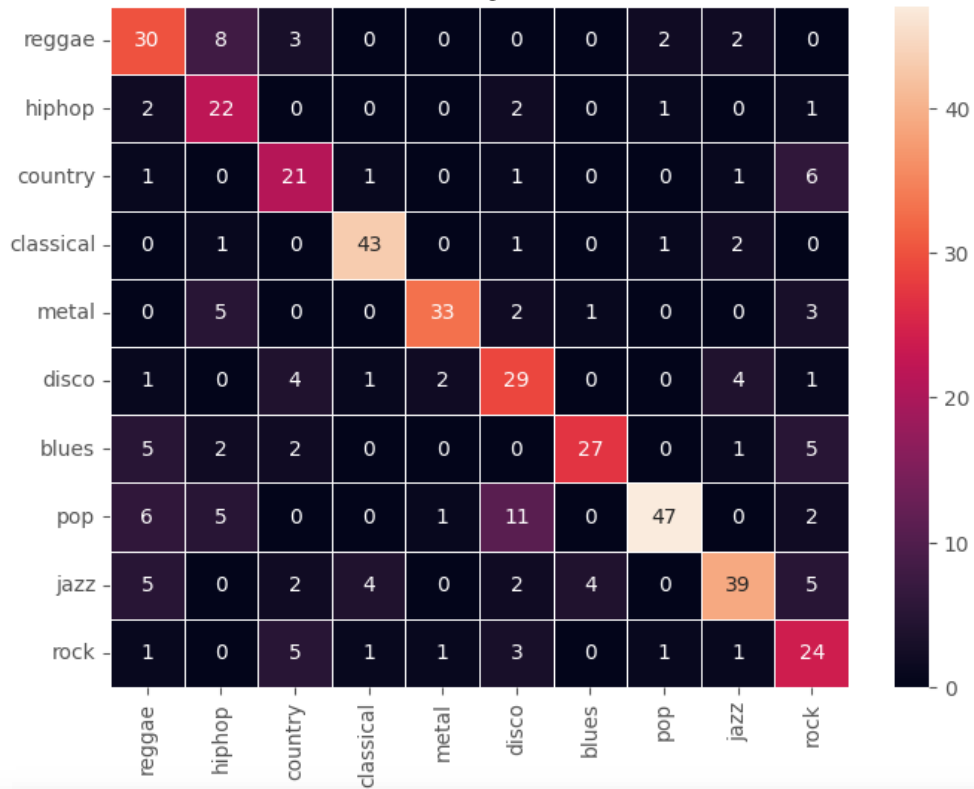
Рекомендации книг

✓ Выводы:

Было очень интересно



Accuracy: 0.70



Анализ сигналов

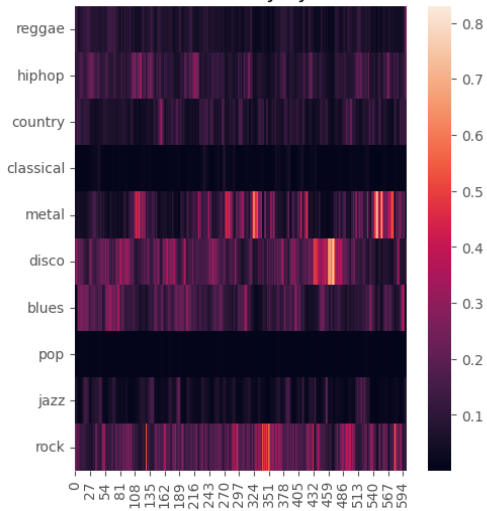
✓ Теория:

Преобразование Фурье,
Short Time Fourier Transform
Мел-шкала, логарифмирование

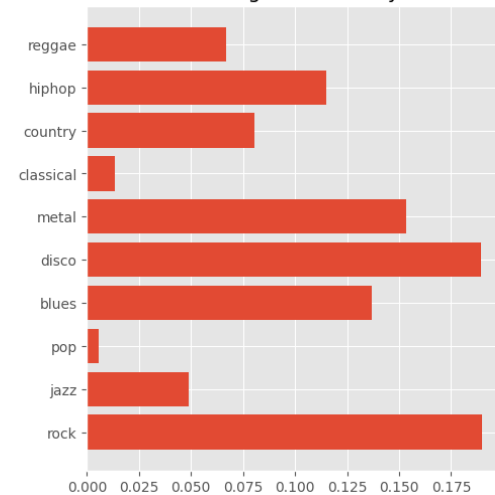
✓ Практика:

Классификация музыкальных жанров

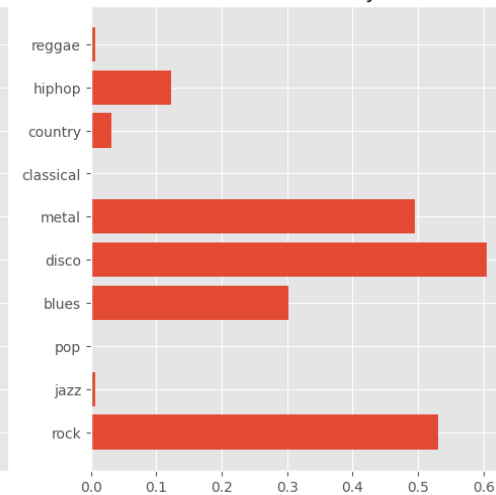
Genre Probability by Frame



Averaged Probability



New Probability



✓ Выводы:

Узнал про преобразование
звуковых данных в вектора

Спасибо за внимание!