

# Классификация музыкальных произведений по жанрам

Илья Шорохов

Финальный отчет по проделанной работе.

## Введение

Сервис "**МиФаСоль**" расширяет работу с новыми артистами и музыкантами, в связи с чем возникла задача – правильно классифицировать новые музыкальные треки, чтобы улучшить работу рекомендательной системы.

Коллеги из отдела работы со звуком подготовили датасет, в котором собраны некоторые характеристики музыкальных произведений и их жанры.

Цель работы – создать модель, позволяющую классифицировать музыкальные произведения по жанрам.

## Данные

В качестве датасета для обучения предоставлен файл **train.csv**, который содержит информацию о примерно 20000 музыкальных треков.

В качестве датасета для тестирования модели предоставлен файл **test.csv**, который содержит информацию примерно о 5000 музыкальных треков.

## Признаки

В наборе данных присутствуют 15 разных признаков, некоторые из них категориальные, другие числовые. Например, признак **key** категориальный и содержит в себе информацию о ключе( тональности) песни. Пример числового признака – **tempo**, который содержит информацию о темпе песни.

# ML-модели, выбранные для тестирования

В качестве моделей для тестирования были выбраны Случайный Лес и Дерево Решений из библиотеки Scikit Learn. Эти модели выбраны так как они лучше линейных моделей справляются с задачей классификации.

## Результаты тестирования моделей

В рамках тестирования моделей случайного леса и дерева решений была выбрана метрика F1-мера.

Для обеих моделей был проведен подбор наиболее оптимальных гиперпараметров. Также перед тестированием была проведена проверка на мультиколлинеарность и удалены признаки, которые не влияли на обучаемость модели.

По итогам тестирования метрика модели случайного леса составила: 0.45

Метрика модели дерева решений: 0.41

В качестве лучшей модели был выбран случайный лес.

## Заключение.

В результате работы была обучена модель случайного леса и получены предсказания для тестовой выборки.

Предсказания были сохранены в отдельный файл с 2 столбцами: id трека и жанром этого трека.