

Проект «Предсказание оценки фильма»

В рамках данного проекта был выбран датасет Netflix TV Shows and Movies, содержащий информацию о фильмах.

Мы объединили два набора данных:

titles.csv — содержит основную информацию о фильмах

credits.csv — содержит информацию об актёрах и их участии в фильмах

После объединения получаем следующие признаки:

1. title — название фильма
2. type — тип (например, фильм, сериал)
3. release_year — год выпуска
4. age_certification — возрастной рейтинг
5. runtime — продолжительность фильма в минутах
6. genres — жанры (один или несколько)
7. production_countries — страна(-ы) производства
8. seasons — количество сезонов (если сериал)
9. imdb_id — идентификатор IMDb
10. imdb_score — оценка IMDb (целевой признак)
11. imdb_votes — количество голосов на IMDb
12. tmdb_popularity — популярность на TMDB
13. tmdb_score — оценка на TMDB
14. actor_name — имя актёра (из файла credits.csv)

Цель проекта

Основная цель — предсказать оценку фильма по версии IMDb на основе различных признаков с помощью модели градиентного бустинга.