

Анализ игр для приставок PS

Автор работы Бондарь Степан

Исходный данные

Данные взяты из датасета PlayStation Games Info, автор Evgeny Baulin

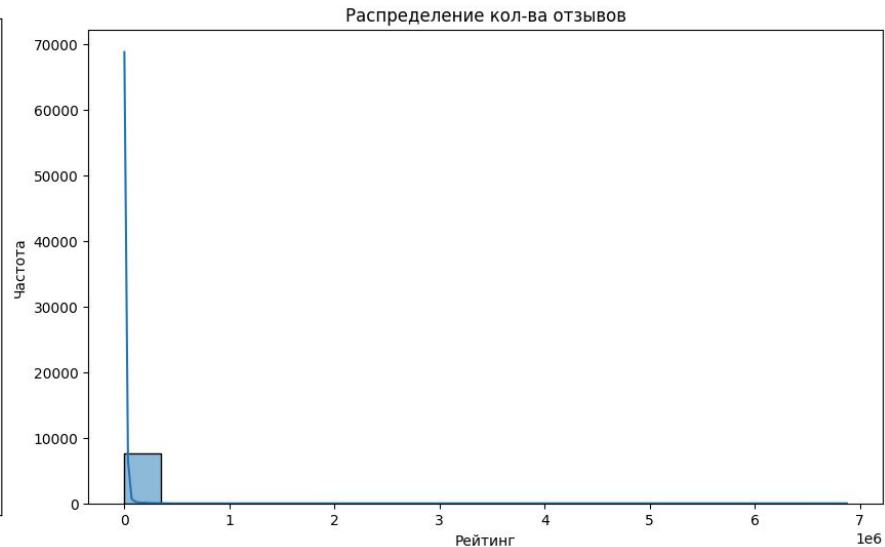
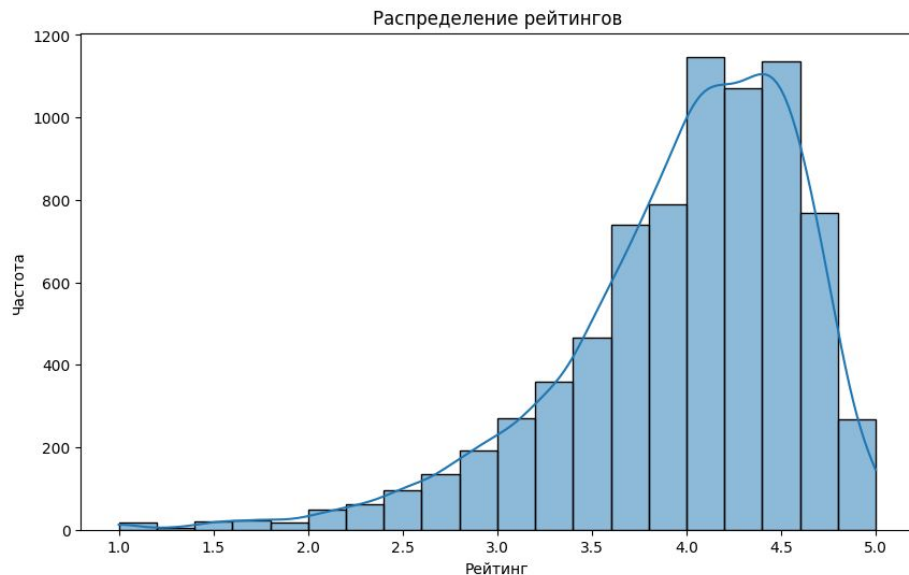
- title - game title. Data in format - string
- platforms - platforms on which the game is available. Data in format - list of strings
- rating - game rating. Data in the format - real number
- votes - number of votes. Data in the format - integer
- developer - game developer. Data in the format - string
- release_date - game release date. Data in the format - date
- price - price of the game in \$. Data in the format - real number
- genres - game genres. Data in format - list of strings
- url - link to the game. Data in format - string

Размер датасета 8981

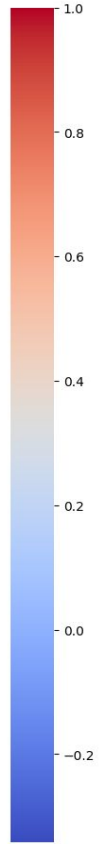
Подготовка данных

- удаляем {'url', 'title'}
- дату преобразуем в год и месяц (в синусной форме)
- с помощью MultiLabelBinarizer преобразуем {'genre', 'platform'} в бинарные признаки
- разделяем на тестовую и тренировочную выборку
- применяем частотное кодирование к 'developer'

Анализ исходных данных



price	1.00	-0.02	-0.02	0.02	-0.10	0.06	0.04	-0.05	0.03	-0.18	-0.00	-0.09	-0.13	0.01	-0.03	0.02	0.04	0.02	0.04	0.05	0.02	-0.14	-0.08	0.19	-0.01	0.02	-0.02	0.03	0.03	0.00	-0.16	0.16	0.01
year	-0.02	1.00	0.17	0.00	-0.25	0.60	-0.15	0.07	0.01	-0.02	-0.05	0.08	0.10	0.00	0.05	0.03	-0.07	0.02	0.02	-0.00	0.02	0.05	0.06	-0.04	-0.06	0.06	0.08	-0.01	-0.03	0.01	0.02	-0.10	-0.06
month_sin	-0.02	0.17	1.00	-0.06	-0.06	0.03	0.01	-0.01	-0.00	-0.01	-0.01	-0.02	0.00	0.00	0.00	-0.00	-0.01	0.02	-0.00	-0.00	-0.01	0.02	0.03	0.02	0.03	-0.03	-0.02	-0.03	-0.00	0.01	0.01	-0.00	-0.01
month_cos	-0.02	0.00	-0.06	1.00	-0.03	0.02	0.03	0.00	0.01	-0.01	-0.01	0.02	-0.00	0.00	0.00	-0.01	-0.00	-0.00	-0.01	0.04	0.01	-0.00	0.03	-0.00	-0.03	0.02	0.01	-0.01	-0.01	0.03	-0.02	-0.01	0.00
PS4	-0.10	-0.25	-0.06	-0.03	1.00	-0.34	0.05	0.02	-0.02	0.03	0.01	0.01	0.01	0.01	-0.01	0.00	0.02	-0.07	-0.03	-0.00	-0.00	-0.01	0.02	0.00	0.00	0.01	0.00	-0.00	-0.00	0.02	0.08	0.06	-0.03
PS5	-0.06	0.60	0.03	0.02	-0.34	1.00	-0.08	0.08	0.04	-0.02	-0.02	0.07	0.06	-0.00	0.02	0.00	-0.05	0.02	0.03	0.01	-0.01	0.04	0.07	-0.01	-0.06	0.03	0.07	-0.02	-0.04	-0.01	0.02	-0.01	0.00
Action	-0.04	-0.15	0.01	-0.03	0.05	-0.08	1.00	0.05	0.01	0.04	-0.02	-0.08	-0.13	-0.08	-0.06	-0.07	0.04	-0.02	0.01	-0.02	-0.04	-0.12	-0.06	-0.04	0.10	-0.13	-0.07	-0.10	-0.09	-0.11	0.09	0.09	0.04
Adult	-0.05	0.07	-0.01	0.00	0.02	0.08	-0.05	1.00	0.06	-0.01	-0.00	0.02	-0.01	-0.02	-0.01	-0.02	-0.01	-0.00	-0.00	-0.01	-0.01	0.05	-0.01	-0.02	-0.03	-0.02	-0.01	-0.02	-0.02	0.01	0.02	0.01	-0.01
Adventure	-0.03	0.01	-0.00	0.01	-0.02	0.04	0.01	0.06	1.00	-0.16	-0.01	0.05	-0.03	-0.10	-0.01	-0.01	0.07	-0.02	0.06	-0.03	-0.07	0.01	-0.05	-0.02	-0.14	-0.08	-0.05	-0.11	-0.10	0.01	-0.08	0.05	0.02
Arcade	-0.18	-0.02	0.01	-0.01	0.03	-0.02	0.04	-0.01	-0.16	1.00	-0.01	-0.03	0.07	0.03	-0.03	-0.04	0.06	-0.00	-0.07	-0.03	0.00	-0.07	-0.03	-0.11	0.10	-0.08	-0.05	0.03	-0.07	-0.06	0.19	-0.01	-0.02
Board Game	-0.00	-0.05	-0.01	-0.01	0.01	-0.02	-0.02	-0.00	-0.01	-0.01	1.00	-0.00	-0.01	-0.00	-0.00	0.02	-0.00	-0.00	-0.00	-0.00	-0.00	-0.01	-0.00	-0.01	-0.01	-0.01	-0.00	-0.00	-0.01	-0.01	0.01	-0.00	-0.00
Brain Training	-0.09	0.08	0.02	0.02	0.01	0.07	-0.08	0.02	-0.05	-0.03	-0.00	1.00	0.04	-0.02	0.08	0.02	-0.02	0.02	-0.02	-0.00	-0.00	0.12	0.31	-0.04	-0.04	-0.03	-0.02	-0.02	-0.01	-0.01	-0.00	-0.01	-0.01
Casual	-0.13	0.10	0.00	-0.00	0.01	0.06	-0.13	-0.01	-0.03	0.07	-0.01	0.04	1.00	-0.03	-0.00	0.21	-0.04	-0.01	-0.06	-0.01	0.08	0.09	-0.02	-0.08	-0.09	0.08	-0.00	-0.01	-0.02	-0.01	0.00	-0.07	-0.02
Driving/Racing	-0.01	0.00	0.00	0.00	-0.01	-0.00	-0.08	-0.02	-0.10	0.0																							



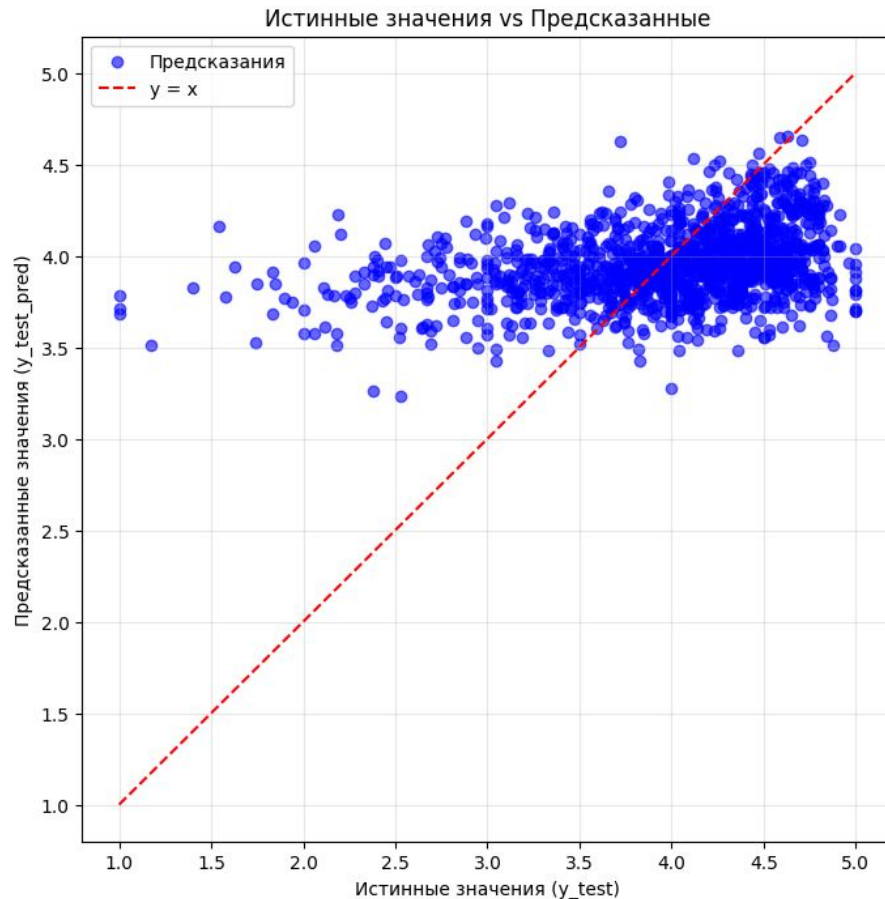
LinearRegression() — rating

Train y (rating) MAPE: 0.1372

Train y (rating) RMSE: 0.6057

Test y (rating) MAPE: 0.1394

Test y (rating) RMSE: 0.6128



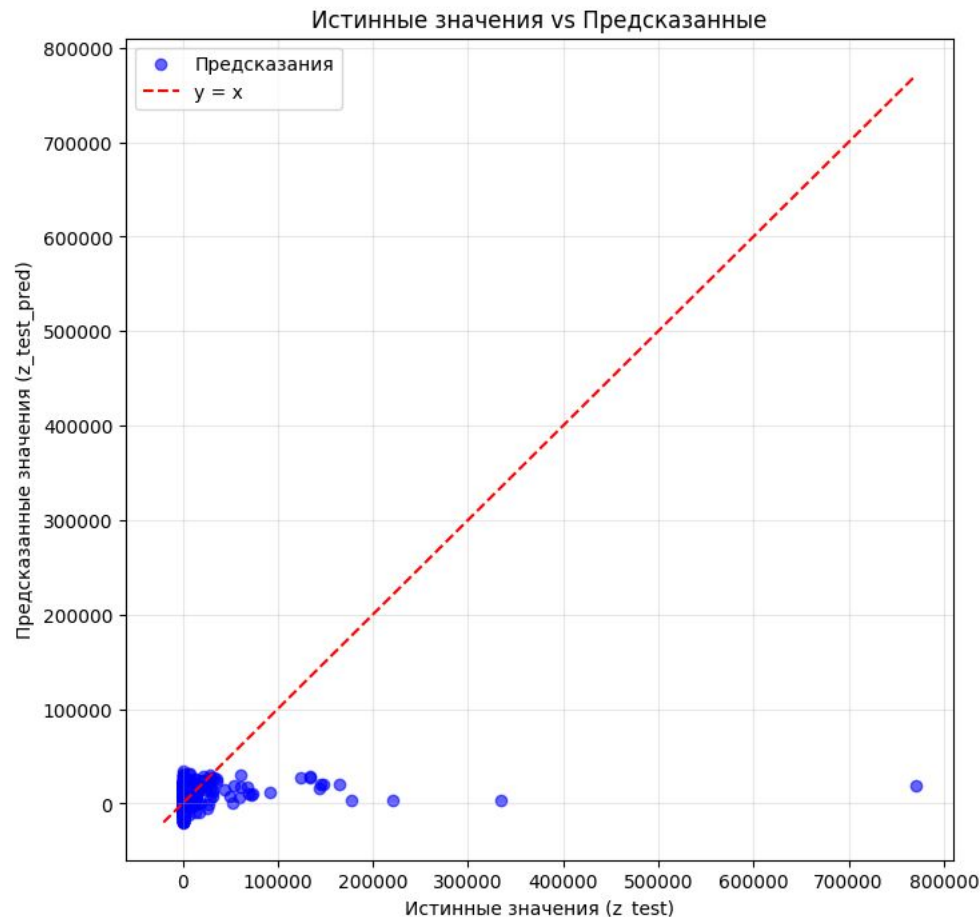
LinearRegression() — votes

Train z (votes) MAPE: 276.1577

Train z (votes) RMSE: 85556.6565

Test z (votes) MAPE: 266.8050

Test z (votes) RMSE: 27398.9408



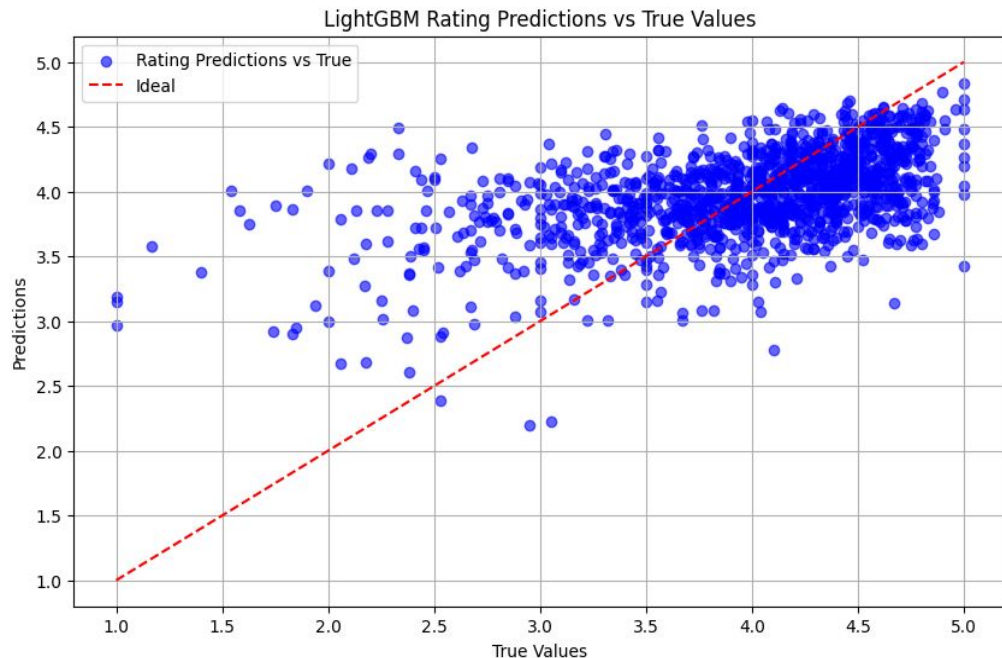
LGBMRegressor() — rating

Train y (rating) MAPE: 0.0964

Train y (rating) RMSE: 0.4388

Train y (rating) MAPE: 0.1251

Train y (rating) RMSE: 0.5619



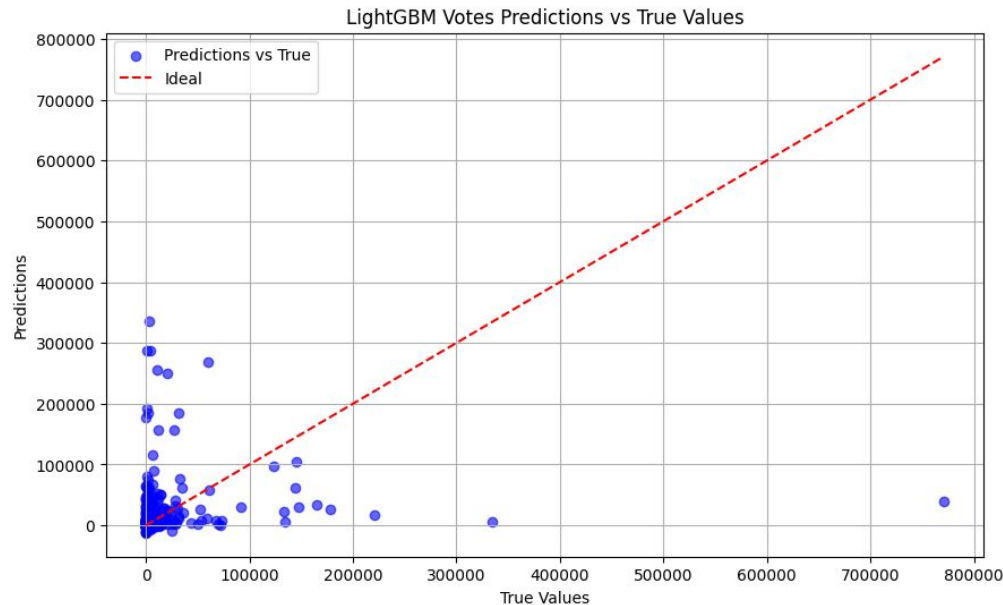
LGBMRegressor() — votes

Train z (votes) MAPE: 56.3689

Train z (votes) RMSE: 78343.3302

Test z (votes) MAPE: 46.4196

Test z (votes) RMSE: 33467.3918



Подбор гиперпараметров

Лучший результат:

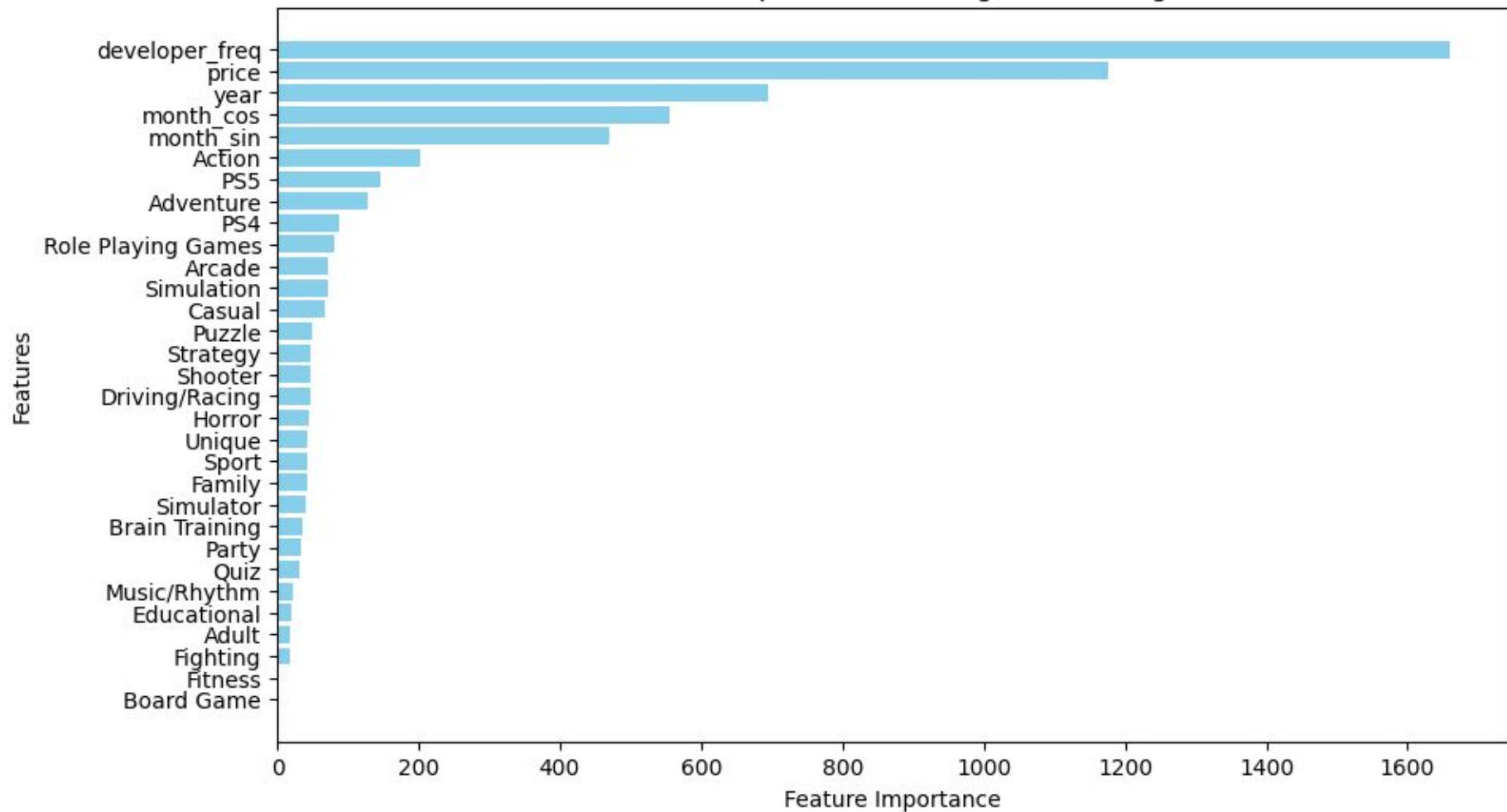
Train y (rating) MAPE: 0.1060

Train y (rating) RMSE: 0.4815

Test y (rating) MAPE: 0.1249

Test y (rating) RMSE: 0.5575

Feature Importance from LightGBM rating



Feature Importance from LightGBM votes

