

Исследование Youtube

Захар Кравчук

21 мая 2021 г.

1 Вступление

Многие из нас пользуются стриминговыми сервисами для просмотра видео. Одним из таких сервисов является Youtube. Недавно из новостей я узнал, что в схожих сервисах администраторы начали скрывать количество "лайков" (положительных оценок о видео) и "дизлайков". В этом проекте я попытаюсь создать модель, способную по мета-информации о видео предсказать количество положительных оценок о нём. Такая модель будет очень полезна на практике, ведь по полученной информации из неё можно судить о качестве видео

2 Постановка задачи

Необходимо обучить модель, которая по мета-информации о видео будет предсказывать количество положительных оценок о видео.

3 Датасет

В своём исследовании я использовал датасет с платформы Kaggle. В нём для каждого видео хранится его название, название канала, на котором оно опубликовано, дата публикации, теги, количество просмотров, количество положительных и отрицательных оценок, количество комментариев и некоторые другие параметры.

4 Эксперименты

Я перебирал параметр регуляризатора в линейной регрессии. По графику видно, что лучше всего взять $\alpha = 10^4$. Также я попробовал использовать решающие деревья, но они показали результат хуже по метрике $RMSE$.

5 Выводы

С помощью решающих деревьев и линейной регрессии я получил две модели, которые обладают достаточно высокой точностью ($RMSE = 76885$). В дальнейшем, в случае если на youtube действительно будут скрыты лайки, можно использовать эти модели.

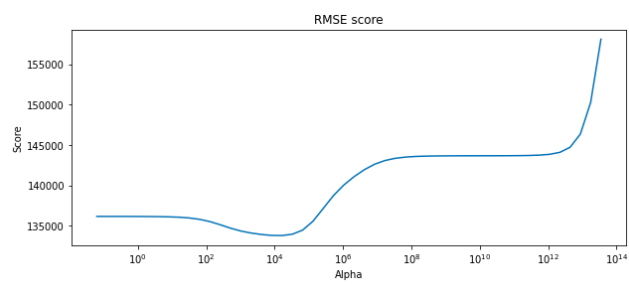


Рис. 1: