

Априорные и апостериорные предикторы востребованности контента

В. Г. Мосин

Аннотация

В статье исследовано влияние априорных признаков контента на точность прогнозирования его востребованности с помощью регрессионной модели. Показано, что использование априорных признаков значительно повышает эффективность моделей по сравнению с моделями, основанными только на апостериорных признаках.

Содержание

1	Введение	1
1.1	Теоретическая часть	2
1.2	Постановки задачи	3
1.2.1	Предмет исследования	3
1.2.2	Методика исследования	3
1.2.3	Цель исследования	3
1.3	Библиотеки	3
2	Описание данных	4
3	Алгоритм	4
3.1	Чтение данных	4
3.2	Разведочный анализ	5
3.3	Модель на числовых данных	6
3.3.1	Удаление строковых данных	6
3.3.2	Построение и обучение регрессионной модели	7
3.3.3	Вывод метрики эффективности	7
3.4	Слепое ранжирование	7
3.4.1	Кодирование предиктора 'Тип видео'	8
3.4.2	Построение, обучение и оценка модели	8
3.5	Ранжирование по среднему значению целевой функции	9
3.5.1	Вычисление рангов	9

3.5.2	Кодирование предиктора 'Тип видео'	10
3.5.3	Построение, обучение и оценка модели	10
3.6	Дробное ранжирование	10
3.6.1	Кодирование предиктора 'Тип видео'	11
3.6.2	Построение, обучение и оценка модели	11
3.7	Кодирование One Hot Encoding	11
3.7.1	Кодирование предиктора 'Тип видео'	11
3.7.2	Построение, обучение и оценка модели	12
4	Результаты	12
5	Выводы	13
6	Литература	14

1 Введение

Все характеристики контента можно разделить на две большие группы по отношению к тому, в какой момент они становятся доступны для исследования: 1) до контакта потребителя с контентом (то есть, на этапе разработки и предварительной подготовки) или 2) после контакта с потребителем (то есть, уже после выхода в эфир). Первую группу мы называем априорными предикторами востребованности, вторую — апостериорными.

Априорные предикторы — это такие характеристики, которые известны о контенте до момента его предъявления или доступа к нему для потребителя. Эти характеристики могут быть определены и учтены еще на этапе создания или размещения контента. Вот некоторые примеры априорных характеристик:

1. Тема контента. Контент может относиться к определенной тематике или жанру, эта характеристика может помочь определить, какие потребители будут в нем заинтересованы.
2. Формат контента. Контент может быть представлен в различных форматах, таких как видео, аудио, текст или графика. Такие априорные характеристики могут указывать на способы взаимодействия с контентом и влиять на предпочтения потребителей.
3. Длительность контента. Эта характеристика может быть важным фактором для потребителей при выборе контента, так как многое зависит от того, насколько много времени они готовы потратить на контент.
4. Уровень сложности. Контент может иметь различные уровни сложности, что относится к интеллектуальному или техническому уровню, требующемуся от потребителя при восприятии контента.

5. Целевая аудитория. В зависимости от характеристик и содержания контента, он может быть оптимизирован и ориентирован на определенную целевую аудиторию.

Напротив, апостериорные предикторы — это характеристики, которые возникают или определяются только после того, как контент был представлен или доступен для потребителя. Эти характеристики могут быть получены из данных о поведении, отзывах, реакциях или взаимодействии потребителей с контентом. Вот некоторые примеры апостериорных характеристик контента:

1. Рейтинг и отзывы. После просмотра или взаимодействия с контентом потребители оставляют свои оценки и отзывы. Это может помочь определить, насколько контент понравился аудитории, и какие мнения о нем сложились.
2. Количество просмотров. Количество просмотров или прослушиваний контента может служить показателем его популярности и привлекательности для аудитории.
3. Время просмотра или взаимодействия. Анализ времени, проведенного пользователями на контенте, или длительности их взаимодействия с ним, может помочь определить, насколько значимым и интересным является контент.
4. Комментарии и обратная связь. Реакции и комментарии пользователей могут дать представление об их отношении к контенту, вызвать вопросы или предложения для дальнейшего развития и улучшения контента.

Апостериорные характеристики контента помогают более полно понять воздействие и эффективность контента после его предоставления. Они могут использоваться для анализа результативности контента, принятия решений и оптимизации его для лучшего удовлетворения требований и интересов аудитории (см. [5], [8]).

1.1 Теоретическая часть

В нашем исследовании мы изучаем востребованность контента, которую мы ассоциируем с суммарным временем его просмотра. Предикторы относятся к апостериорному классу, и, добавляя еще один априорный предиктор, мы сравниваем результаты. Связь востребованности с предикторами мы описываем при помощи нескольких регрессионных моделей, каждая из которых обладает определенной спецификой.

1. Модель на апостериорных предикторах. Это регрессионная модель на числовых данных, она не требует предварительной обработки данных.

2. Смешанная модель на апостериорных предикторах с добавлением одного категориального априорного предиктора. Здесь требуется предварительное кодирование категорий, для чего могут использоваться различные методы.

Методы кодирования категориальных данных хорошо изучены и повсеместно применяются на практике. Однако нет единого универсального ответа, какой из них оказывается наиболее эффективным в той или иной ситуации. Поэтому подбор метода происходит эмпирически, путем прямого сравнения полученных результатов.

1.2 Постановки задачи

1.2.1 Предмет исследования

Анализируются данные о видеороликах, расположенных на одном из ведущих хостингов. Данные являются суммарными и относятся к годовому промежутку с 01.01.2022 по 01.01.2023.

1.2.2 Методика исследования

Методика состоит в построении регрессионной модели, описывающей зависимость целевой функции от предикторов и вычислении метрики ее эффективности. Модель строится в двух ситуациях: 1) без использования априорного предиктора, 2) с добавлением априорного предиктора.

1.2.3 Цель исследования

Наша цель состоит в том, чтобы выяснить оказывает ли добавление апостериорного предиктора положительное влияние на качество регрессионной модели, и если да, убедиться в том, что это влияние не случайно.

1.3 Библиотеки

В своих вычислениях мы используем среду `Jupyter Notebook` для языка `Python` и его библиотек: `numpy`, `pandas`, `sklearn`.

Одной из ключевых библиотек для научных вычислений в `Python` является библиотека `numpy`. Она обеспечивает эффективные алгоритмы, функции и структуры данных для работы с числовыми массивами различных размерностей (см. [2], [3], [4]).

Библиотека `pandas` — это мощный инструмент для работы с данными в `Python`, пользующийся большой популярностью. Она предлагает широкий спектр функций для анализа, обработки и манипуляции с данными, включая индексирование, фильтрацию, сортировку, агрегацию, слияние и другие операции (см. [1]).

В Python доступна библиотека `scikit-learn` (она же известна как `sklearn`), которая включает в себя множество инструментов и функций для работы с задачами классификации, регрессии, кластеризации, понижения размерности, выбора моделей и прочего. Мы применяем эту библиотеку для решения регрессионных задач, возникающих при аппроксимации (см. [2], [4]).

2 Описание данных

Фрейм содержит записи о 500 роликах, каждый ролик описан при помощи 20 предикторов и 1 целевого параметра, в качестве которого выступает суммарное время просмотра ролика за период с 01.01.2022 по 01.01.2023. Предикторы относятся к двум классам. 18 предикторов являются апостериорными ('Среднее число просмотров одним пользователем', 'Клики по элементам конечной заставки' и т. д.), 2 предиктора являются априорными, это 'Название видео' и 'Тип видео'.

3 Алгоритм

В этом разделе мы построим и обучим несколько регрессионных моделей. Сначала мы вообще не будем использовать сведения об априорных характеристиках объектов, а затем включим их в модели, применяя различные способы их кодирования в числовой формат.

Начнем с того, что прочитаем данные и сформируем о них какое-то первичное представление.

3.1 Чтение данных

Применяя метод `read_csv` библиотеки `pandas`, формируем дата-фрейм:

	Название видео	Тип видео	Поделиться	...	Время просмотра
0	Линейная зависимость и линейная...	Тема	27	...	90.9871
1	Вычисление определителей по методу...	Тема	59	...	49.8394
2	Приведение кривой второго порядка...	Пример	41	...	38.4302
...
498	Градиент. Вопросы	Вопросы	2	...	0.3030

	Название видео	Тип видео	Поделиться	...	Время просмотра
499	Криволинейный интеграл второго рода...	Ответы	5	...	0.3026

Каждая строка описывает определенный видеоролик, каждый столбец определяет один из признаков, при помощи которого этот ролик описывается. Целевая функция — 'Время просмотра (часы)'.

3.2 Разведочный анализ

Применяя метод `describe` библиотеки `pandas`, выводим сведения о числовых признаках:

	mean	std	min	max
Поделиться	0.640000	1.811343	0.00000	17.000000
Клики по элементам конечной заставки	0.508000	1.048874	0.00000	9.000000
Показы элементов конечной заставки	58.78600	79.56626	0.00000	792.000000
Показы тизеров	3.942000	7.925419	0.00000	56.000000
Постоянные зрители	45.594000	49.457591	2.00000	338.000000
Новые зрители	32.586000	94.640204	0.00000	1100.000000
Среднее число просмотров одним пользователем	1.241784	0.171804	1.00000	2.600000
Уникальные зрители	79.674000	133.822347	3.00000	1465.000000
Средний процент просмотра	35.109640	13.616501	9.520000	74.570000
Отказались от подписки	0.010000	0.099598	0.000000	1.000000
Новые подписчики	0.3600	1.0002	0.000000	9.000000
Новые комментарии	0.056000	0.254939	0.000000	2.000000
Отметки "Не нравится"	0.064000	0.322664	0.000000	4.000000
Отметки "Нравится"	1.486000	3.079085	0.000000	40.000000
Просмотры	100.530000	170.646446	5.000000	1880.000000
Подписчики	0.350000	0.976426	0.000000	9.000000
Показы	621.738000	797.378602	39.000000	8889.000000
CTR для значков видео (%)	5.969620	3.318488	0.000000	21.240000
Время просмотра (часы)	4.023387	7.663491	0.302600	90.987100

Характеристики строковых признаков таковы:

	unique	top	freq
Название видео	500	Линейная зависимость и линейная независимость...	1
Тип видео	5	Тема	204

Значение предиктора 'Название видео' уникально для каждого ролика. Применяя к предиктору 'Тип видео' метод `unique` библиотеки `pandas`, выводим все его уникальные значения. Получаем пять категорий: 'Тема', 'Вопросы', 'Ответы', 'Пример', 'Другое'.

3.3 Модель на числовых данных

Сначала мы решаем задачу в простейшем варианте, используя в качестве предикторов только числовые значения.

3.3.1 Удаление строковых данных

Применяем метод `drop` библиотеки `pandas` и удаляем признаки 'Название видео' и 'Тип видео'. Получаем дата-фрейм, содержащий 500 строк и 18 столбцов.

	Поделиться	Клики по элементам...	Показы элементов...	...	Время просмотра
0	17.0	9.0	27	...	90.9871
1	10.0	7.0	59	...	49.8394
2	8.0	1.0	41	...	38.4302
...
498	0.0	0.0	2	...	0.3030
499	0.0	0.0	5	...	0.3026

Сведений о названии видео и его типе в этом наборе данных нет вообще. Зато все данные стали числовыми, и к ним можно применять процедуры построения и обучения модели в библиотеке `sklearn`.

3.3.2 Построение и обучение регрессионной модели

Для получения левой части регрессионной задачи мы, пользуясь методом `drop` библиотеки `pandas`, удаляем целевой признак 'Время про-

смотря (часы)’ и переводим оставшийся дата-фрейм в массив `numpy`, пользуясь методом `to_numpy`. В результате получается числовой массив `X`, содержащий 500 строк и 17 столбцов. Для получения правой части выделяем целевой признак, переводим его в массив и получаем одномерный числовой массив `y`, содержащий 500 элементов. Затем, методом `LinearRegression` из модуля `linear_model` библиотеки `sklearn`, мы формируем объект `model` и применяем к нему метод `fit` на `X` и `y`.

3.3.3 Вывод метрики эффективности

Применяя к обученной модели `model` метод `score`, получаем значение коэффициента детерминации: $R^2 = 0.9134$. Это очень хорошее значение для метрики эффективности регрессионной модели. Но возникает естественный вопрос, нельзя ли сделать этот показатель еще лучше, ведь при построении модели мы не учитывали значение категориального признака ‘Тип видео’.

3.4 Слепое ранжирование

В исходном наборе данных присутствуют два нечисловых предиктора: ‘Название видео’ и ‘Тип видео’. Название уникально для каждого ролика, то есть, является своеобразным ID объекта и абсолютно не информативно с точки зрения прогнозирования целевой функции. Поэтому признак ‘Название видео’ мы, как и выше, просто удаляем.

Предиктор ‘Тип видео’ относит каждый ролик к одной из пяти возможных категорий: ‘Тема’, ‘Вопросы’, ‘Ответы’, ‘Пример’, ‘Другое’ — и здесь следует упомянуть об одной особенности категориальных данных. Дело в том, что категориальные данные могут обладать естественным порядком (например, значения воинских званий ‘Лейтенант’, ‘Старший лейтенант’, ‘Капитан’, ‘Майор’ упорядочены по возрастанию), но могут и не обладать им. Если порядок есть, то целочисленное ранжирование трудно назвать слепым: меньшему по порядку значению присваивается меньший ранг, а большему — больший.

Но в нашем случае порядка нет. Нет никаких оснований утверждать, что категория ‘Пример’ находится выше категории ‘Вопросы’ и т. д. Поэтому значениям категорий признака ‘Тип видео’ мы присваиваем слепые номера: ‘Тема’ — 0, ‘Вопросы’ — 1, ‘Ответы’ — 2, ‘Пример’ — 3, ‘Другое’ — 4. На уровне логики все получается абсолютно примитивно, можно сказать, мы расставили ранги случайно. На уровне технологической реализации нам нужно заменить в нашем дата-фрейме строковые значения категорий на из числовые коды (эта операция называется кодированием).

3.4.1 Кодирование предиктора 'Тип видео'

Применяя метод `replace` библиотеки `pandas`, заменяем строковые значения признака 'Тип видео' их номерами. Получаем дата-фрейм, состоящий из 500 строк и 19 столбцов.

	Тип видео	Поделиться	Клики по элементам...	...	Время просмотра
0	0	17.0	9.0	...	90.9871
1	0	10.0	7.0	...	49.8394
2	3	8.0	1.0	...	38.4302
...
498	1	0.0	0.0	...	0.3030
499	0	0.0	0.0	...	0.3026

По сравнению с предыдущим шагом появился еще один предиктор, причем, он тоже относится к числовому типу, и теперь к данным можно применять процедуры построения и обучения модели в библиотеке `sklearn`.

3.4.2 Построение, обучение и оценка модели

Так же как и выше, строим и обучаем регрессионную модель, после чего вычисляем ее коэффициент детерминации: $R^2 = 0.9205$. Значение метрики эффективности увеличилось во втором знаке после запятой по сравнению с моделью на числовых данных.

3.5 Ранжирование по среднему значению целевой функции

Недостатки слепого целочисленного ранжирования очевидны: мы произвольным образом упорядочили метки типа, и если занумеровать их в другом порядке, то результат моделирования целевой функции тоже будет другим (хотя, скорее всего, не будет принципиально отличаться от полученного выше). Другой подход к ранжированию категорий состоит в использовании информации о поведении целевой функции. Следует понимать, что для той или иной категории объектов категориального типа среднее значение целевой функции, во-первых, почти наверняка не совпадает со средними значениями целевой функции на других категориях, а во-вторых, что это несовпадение можно и нужно использовать для построения более точной прогнозирующей модели.

3.5.1 Вычисление рангов

При помощи метода `loc` библиотеки `pandas` мы формируем локализации набора данных по условию равенства предиктора 'Тип видео' тому или иному значению. Получаем 5 локализаций, для каждой из которых, пользуясь методом `mean`, вычисляем среднее значение целевой функции 'Время просмотра (часы)'.

Значение типа	Среднее значение целевой функции на локализации	Номер слепого ранжирования
Ответы	1.448203100775193	2
Другое	1.0165177777777777	4
Вопросы	0.783105952380952	1
Пример	5.612536842105264	3
Тема	7.353307352941177	0

На множестве средних значений (в отличие от множества типов видео) есть естественное отношение порядка. Теперь мы можем упорядочить типы видео по среднему значению целевой функции на соответствующей локализации и получить новые номера рангов.

Значение типа	Среднее значение целевой функции на локализации	Номер ранга по среднему значению
Вопросы	0.783105952380952	0
Другое	1.0165177777777777	1
Ответы	1.448203100775193	2
Пример	5.612536842105264	3
Тема	7.353307352941177	4

3.5.2 Кодирование предиктора 'Тип видео'

Применяя метод `replace` библиотеки `pandas`, мы заменяем строковые значения признака 'Тип видео' их номерами, но на этот раз используем не слепые номера, а те, что были получены по среднему значению целевой функции на шаге 3.5.1. Получаем дата-фрейм, состоящий из 500 строк и 19 столбцов.

	Тип видео	Поделились	Клики по элементам...	...	Время просмотра
0	4	17.0	9.0	...	90.9871
1	4	10.0	7.0	...	49.8394
2	3	8.0	1.0	...	38.4302
...
498	0	0.0	0.0	...	0.3030
499	4	0.0	0.0	...	0.3026

3.5.3 Построение, обучение и оценка модели

Так же как и выше, строим и обучаем регрессионную модель, после чего вычисляем ее коэффициент детерминации: $R^2 = 0.9306$. Значение метрики эффективности увеличилось во втором знаке после запятой как по сравнению с моделью на числовых данных, так и с моделью со слепым кодированием.

3.6 Дробное ранжирование

Ориентируясь на значения целевой функции, можно предложить еще один вариант ранжирования: использовать в качестве рангов не номера (то есть, неотрицательные целые числа, следующие друг за другом без пропусков), а непосредственно средние значения целевой функции, вычисленные на соответствующей локализации данных (то есть, дробные числа).

3.6.1 Кодирование предиктора 'Тип видео'

Средние значения целевой функции на локализациях мы получили выше. Теперь, пользуясь методом `replace` библиотеки `pandas`, мы заменяем строковые значения признака 'Тип видео' найденными средними значениями.

	Тип видео	Поделились	Клики по элементам...	...	Время просмотра
0	7.353307352	17.0	9.0	...	90.9871
1	7.353307352	10.0	7.0	...	49.8394
2	5.612536842	8.0	1.0	...	38.4302
...
498	0.783105952	0.0	0.0	...	0.3030

	Тип видео	Поделиться	Клики по элементам...	...	Время просмотра
499	7.353307352	0.0	0.0	...	0.3026

В столбце 'Тип видео' по-прежнему находится 5 различных значений, но теперь это не строки и не целые числа, а числа с плавающей запятой.

3.6.2 Построение, обучение и оценка модели

Так же как и выше, строим и обучаем регрессионную модель, после чего вычисляем ее коэффициент детерминации: $R^2 = 0.9297$. Здесь мы впервые наблюдаем неоднозначный эффект. С одной стороны, значение метрики эффективности увеличилось во втором знаке после запятой по сравнению с моделью на числовых данных и моделью со слепым ранжированием строкового признака 'Тип видео'. С другой стороны, оно незначительно, но все-таки уменьшилось относительно модели с целочисленным ранжированием по средним значениям целевой функции.

3.7 Кодирование One Hot Encoding

Этот метод кодирования категориальных данных не является методом ранжирования и вообще не апеллирует к понятию отношения порядка на множестве. Его идея состоит в том, чтобы кодировать категориальное значение не числом, а вектором, все координаты которого равны 0, кроме одной, которая соответствует определенной категории и равна 1.

3.7.1 Кодирование предиктора 'Тип видео'

Применяя метод `get_dummies` библиотеки `pandas`, переводим каждое значение предиктора 'Тип видео' в пятимерный вектор.

	x_1	x_2	x_3	x_4	x_5	Поделиться	...	Время просмотра
0	1	0	0	0	0	27	...	90.9871
1	1	0	0	0	0	59	...	49.8394
2	0	0	0	1	0	41	...	38.4302
...
498	0	1	0	0	0	2	...	0.3030
499	1	0	0	0	0	5	...	0.3026

Здесь маркер x_i означает принадлежность объекта к типу 'Тема', 'Вопросы', 'Ответы', 'Пример' и 'Другое' соответственно.

В этой таблице 24 столбца, из которых 19 столбцов — это все имевшиеся с самого начала апостериорные признаки (включая целевую функцию 'Время просмотра (часы)'), и еще 5 столбцов — это искусственно созданные столбцы на основании данных априорного строкового признака 'Тип видео'.

3.7.2 Построение, обучение и оценка модели

Так же как и выше, строим и обучаем регрессионную модель, после чего вычисляем ее коэффициент детерминации: $R^2 = 0.9345$. Мы видим, что кодирование по методу One Hot Encoding превосходит другие методы кодирования по метрике эффективности.

4 Результаты

Были рассмотрены данные о 500 роликах, описанных при помощи 20 предикторов и 1 целевого параметра, в качестве которого выступает суммарное время просмотра ролика за период с 01.01.2022 по 01.01.2023, причем, 18 предикторов являются апостериорными ('Среднее число просмотров одним пользователем', 'Клики по элементам конечной заставки' и т. д.), а 2 предиктора являются априорными, это 'Название видео' и 'Тип видео'.

На основании этих данных были построены регрессионные модели, прогнозирующие время просмотра по набору апостериорных признаков и одному априорному признаку 'Тип видео'. Одна из моделей вообще не использовала априорный категориальный признак, остальные модели включали его в качестве дополнительного предиктора, причем, использовали разные методы кодирования категорий в числовой формат. В качестве метрики эффективности регрессионных моделей использовался коэффициент детерминации R^2 . Для моделей, рассмотренных в настоящем исследовании, получены следующие значения этой метрики:

Модель без использования априорного предиктора	$R^2 = 0.9134$
Модели с использованием априорного предиктора	
Слепое ранжирование	$R^2 = 0.9205$
Ранжирование по среднему значению целевой функции	$R^2 = 0.9306$
Дробное ранжирование	$R^2 = 0.9297$
One Hot Encoding	$R^2 = 0.9345$

Наиболее эффективным оказалось использование априорного признака с переводом его строковых значений в числовые векторы методом One Hot Encoding.

5 Выводы

В исследовании регрессионной зависимости востребованности контента от различных предикторов важную роль играют априорные, то есть, известные до контакта с потребителем, признаки.

Это связано с тем, что прогноз востребованности контента на основе его априорных признаков позволяет принимать более обоснованные решения по созданию, продвижению и маркетингу контента, что повышает его эффективность, удовлетворяет потребности аудитории и улучшает результативность контентной стратегии. Такой прогноз поможет сфокусировать маркетинговые усилия на наиболее перспективных контентных предложениях, что позволит увеличить эффективность и результативность маркетинговых кампаний. Зная, какой контент имеет больший шанс быть популярным, можно нацелиться на создание контента, который лучше удовлетворит интересы и потребности целевой аудитории, повысив их удовлетворенность и лояльность. И самое главное, прогноз востребованности контента по его априорным признакам позволяет оптимизировать бюджет, направляя ресурсы на разработку контента, который наиболее вероятно будет успешным, и сокращая расходы на менее перспективные контентные проекты (см. [6], [7]).

К сожалению, получение априорных признаков востребованности контента может быть очень сложной задачей. Во-первых, для создания априорных признаков востребованности необходимо иметь доступ к большому объему данных, включая информацию о прошлых результатах, характеристики целевой аудитории, особенности контента и т. д. Однако иногда эти данные могут быть недоступны или неполны. Во-вторых, априорные признаки востребованности контента могут быть сложными для моделирования из-за их многообразия и динамичности. Например, учитывая изменчивость интересов аудитории, поведенческие особенности и воздействие конкурентов, создание точной и стабильной модели предсказания востребованности контента может оказаться сложным. И наконец, некоторые априорные признаки, такие как уникальность и качество контента, могут быть субъективными и неподдающимися количественному измерению (см. [6], [7]).

Вместе с тем, получить апостериорные признаки востребованности контента может быть относительно легко, благодаря наличию большого объема данных, легкости в получении информации из различных источников и доступности инструментов аналитики. С развитием цифровых платформ, таких как социальные сети, хостинги, блоги и т. д., произво-

дители контента имеют доступ к множеству данных о поведении своей аудитории, включая информацию о просмотрах, лайках, комментариях, репостах, а также данные из систем аналитики, таких как Google Analytics (см. [5], [8]).

Однако следует учитывать, что использование исключительно апостериорных признаков для прогнозирования востребованности контента имеет ряд ограничений, которые могут сделать их недостаточными для решения большинства маркетинговых задач. Например, апостериорные признаки могут указывать на популярность контента в прошлом, но они не всегда отражают реальную ценность или качество контента. Контент может быть востребованным из-за различных факторов, включая вирусный характер, специфические тренды или реакции аудитории, что не всегда отображается просто в количестве просмотров или лайков. Или апостериорные признаки могут указывать на реакцию аудитории, но они не всегда предоставляют глубокого понимания внутренних мотиваций и потребностей аудитории.

Для успешного маркетинга необходимо понимать, что именно движет целевой аудиторией, что они ценят, и какие проблемы они хотят решить.

6 Литература

1. Хейдт М. Изучаем Pandas / М. Хейдт — Москва: ДМК Пресс, 2018. — 438 с.
2. Бурков А. Машинное обучение без лишних слов / А. Бурков — СПб: Питер, 2020. — 192 с.
3. Вьюгин, В. В. Математические основы теории машинного обучения и прогнозирования / В. В. Вьюгин — М.: МЦИМО, 2013. — 387 с.
4. Бринк Х. Машинное обучение / Х. Бринк, Дж. Ричардс, М. Феве-ролф — СПб.: Питер, 2017. — 336 с.
5. Гусков И.В. Предикторы востребованности контента в социальных сетях // Вестник Московского университета. Серия 10: Журналистика. 2015. Т. 1. С. 68–81.
6. Мещерякова Е.Ю., Морозова О.В. Предикторы эмоциональной реакции на контент в социальных сетях // Социальные коммуникации: теория и практика. 2019. Т. 2. № 1. С. 68–79.
7. Савин А.В. Психологические предикторы эффективности контента в онлайн-среде // Вестник Тюменского государственного университета. 2018. № 6. С. 77–83.
8. Чижикова О.В., Федотова О.И. Анализ предикторов востребованности видеоконтента в социальных сетях // Информационные технологии и вычислительные системы. 2018. № 2(89). С. 56–61.