

---

# Определение тональности отзывов при помощи машинного обучения

Жукова Екатерина Сергеевна,  
ООО «Энавер Поинт»

Социальные сети и сервисы постоянно развиваются, и все больше людей могут выразить свое мнение о товарах, услугах, фильмах, других людях публично. Чтобы определить отношение аудитории к организации или товару, приходится анализировать десятки тысяч сообщений. Обращивать такой объем информации вручную становится невозможным.

В статье представлены результаты исследования по использованию машинного обучения для разметки тональности отзывов в рамках технологии «Vabkee».

## Подготовка данных для обучения

Изначально перед нами стояла задача классификации на три класса.

В качестве источника данных брали сайты, на которых люди оставляют отзыв и присваивают ему баллы: например, 1 — если они не довольны товаром или услугой, 5 — если очень довольны. Мы выбрали сайт [otzovik.com](http://otzovik.com), на нем публикуют различные отзывы, в том числе о компаниях и услугах, и есть шкала удовлетворенности от 1 до 5. С сайта были собраны более 100 тысяч отзывов.

Хорошим источником мог стать Яндекс.Маркет. Но от него отказались, так как он в большей степени посвящен товарам и имеет платный API.

## Подготовка текстов отзывов к обучению

Все тексты были приведены к нижнему регистру и нормализованы, также были удалены символы пунктуации. Очистка текста от стоп-слов и служебных частей речи не проводилась, так как текст «Мне нравится» и «Мне не нравится» имеет совершенно разную эмоциональную окраску, которая стирается без частицы не.

## Обучение

Для начала тексты необходимо представить в виде вектора чисел. Сделать это можно разными способами. Самый простой — это мешок слов, модель текстов на натуральном языке, в которой каждый документ или текст выглядит как неупорядоченный набор слов без сведений о связях между ними. В библиотеке `sklearn` это метод `CountVectorizer`.

Более сложный вариант, учитывающий частотность слов, — TF-IDF. Это статистическая мера, используемая для оценки важности слова в контексте [документа](#), являющегося частью коллекции документов или [корпуса](#). Большой вес в TF-IDF получают слова с высокой [частотой](#) в пределах конкретного документа и с низкой частотой употреблений в других документах:

---

$$\text{tf-idf}(t, d, D) = \text{tf}(t, d) \times \text{idf}(t, D) ,$$

$$\text{tf}(t, d) = \frac{n_t}{\sum_k n_k}$$

где

$$\text{idf}(t, D) = \log \frac{|D|}{|\{d_i \in D \mid t \in d_i\}|}$$

Мы использовали именно его. В библиотеке sklearn это метод TfidfVectorizer.

Как ранее упоминалось, нам нужно уметь различать «нравится» и «не нравится», для этого необходимо в словарь слов, по которому мы будем составлять вектор, добавить не только одиночные слова, но и частые биграммы (словосочетания). Это можно сделать средствами sklearn, передав методу TfidfVectorizer параметр ngram\_range.

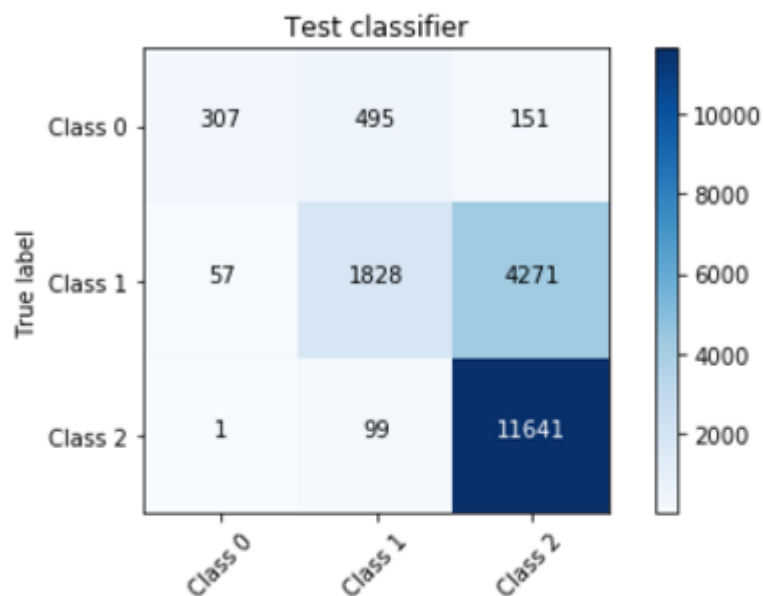
Само обучение проводилось с помощью стохастического градиентного спуска (Stochastic Gradient Descent), реализация которого есть в sklearn. Этот метод хорошо работает с задачами, где присутствуют разреженные данные. Как раз такой задачей и является классификация текстов.

После обучения каждой модели, мы получили прогноз на валидационной выборке, построили матрицу неточности и отчет классификации.

К первому классу (Class 0) относились отзывы с отметкой 1, ко 2 классу (Class 1) — с отметками 2-4, к 3 классу (Class 2) — с отметкой 5.

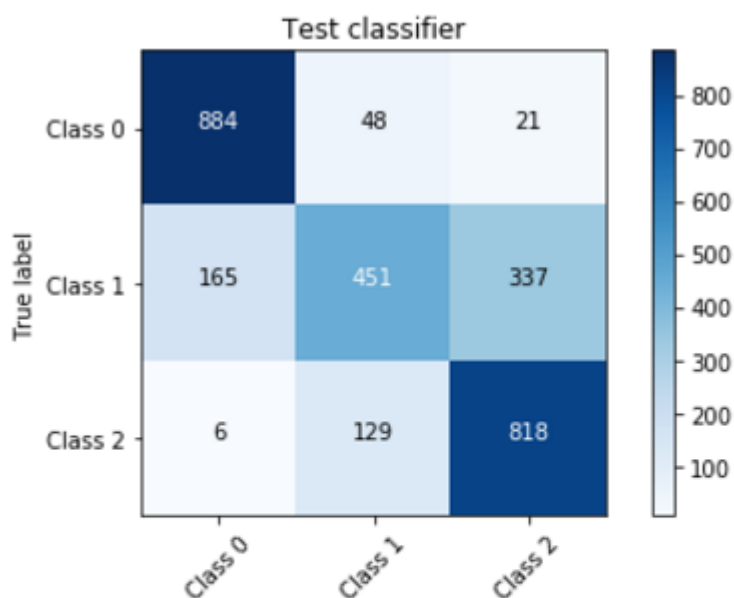
Точность (precision) в пределах класса — это доля объектов, действительно принадлежащих данному классу относительно всех объектов, которые система отнесла к этому классу.

Полнота (recall) класса — это доля найденных классификатором объектов, принадлежащих классу относительно всех объектов этого класса в тестовой выборке.



	precision	recall	f1-score	support
Class 0	0.84	0.32	0.47	953
Class 1	0.75	0.30	0.43	6156
Class 2	0.72	0.99	0.84	11741
avg / total	0.74	0.73	0.68	18850

Точность всех классов достаточно неплохая — в среднем 74%, что говорит о том, что основные слова-маркеры были найдены. Но есть сильное смещение в пользу Class 2. Чтобы его убрать, мы уравнили число объектов в каждом классе.



	precision	recall	f1-score	support
Class 0	0.84	0.93	0.88	953
Class 1	0.72	0.47	0.57	953
Class 2	0.70	0.86	0.77	953
avg / total	0.75	0.75	0.74	2859

Ситуация улучшилась, но нейтральный класс (Class 1) имеет очень низкую полноту, то есть отзывы этого класса классификатор в 50% случаев пытается определить либо как позитивные, либо как негативные. Класс позитивных отзывов также не очень хорошо работает. После анализа ошибок классификатора было выяснено, что так происходит потому, что отзывы с отметкой 4 очень сложно отличить от отзывов с отметкой 5.

Решить проблему с нейтральными отзывами, удалив из обучения все сообщения с отметкой 4, не получится, потому что при классификации на три класса каждый отзыв может иметь только один вид тональности. Однако существуют отзывы, которые имеют и положительную и отрицательную тональность. Например:

Достоинства: действительно чистит, экономичное Недостатки: перехимичили, противно резко пахнет, стойкий запах Здравствуйте, приобрела \*АОС я родилась с первых дней жизни с ромашкой\* для мытья посуды по привлекательной цене и разочаровалась. С момента открытия бутылочки обнаружила неприятный запах напоминающий краску или растворитель

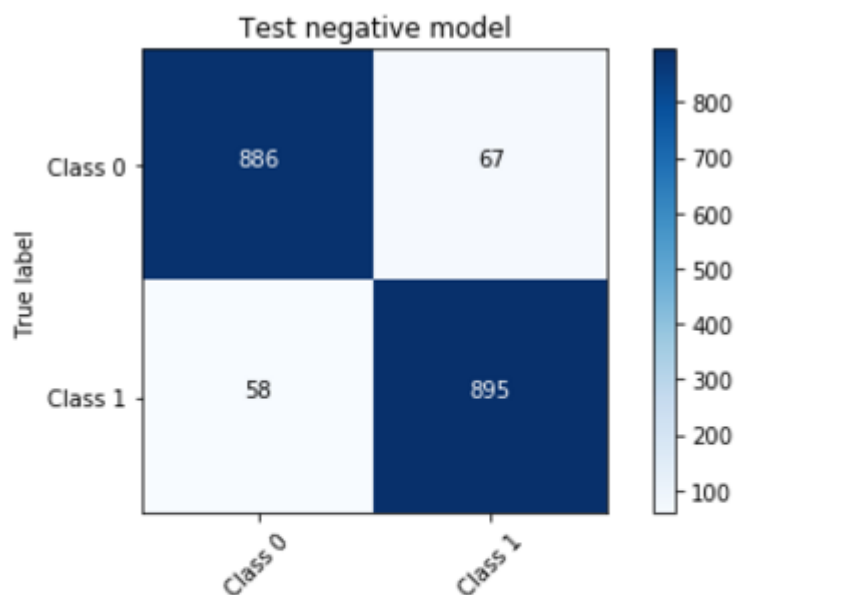
Чтобы решить эту проблему, мы создали две модели:

- бинарная классификация: негативная тональность у отзыва или нет;
- бинарная классификация: положительная тональность у отзыва или нет.

После обучения каждой модели на валидационной выборке мы получили прогноз и построили матрицу неточности и отчет классификации.

### Результаты обучения модели с негативной тональностью

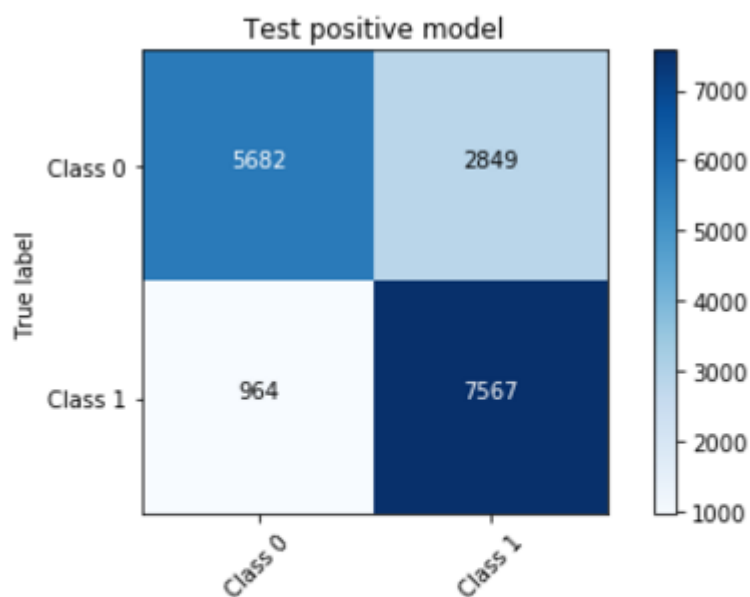
В валидационной выборке — 1906 сообщений. Class 0 — сообщения, в которых нет негативной тональности. Class 1 — сообщения, в которых есть негативная тональность. Полнота и точность результатов высокая — в среднем 93%.



	precision	recall	f1-score	support
Class 0	0.94	0.93	0.93	953
Class 1	0.93	0.94	0.93	953
avg / total	0.93	0.93	0.93	1906

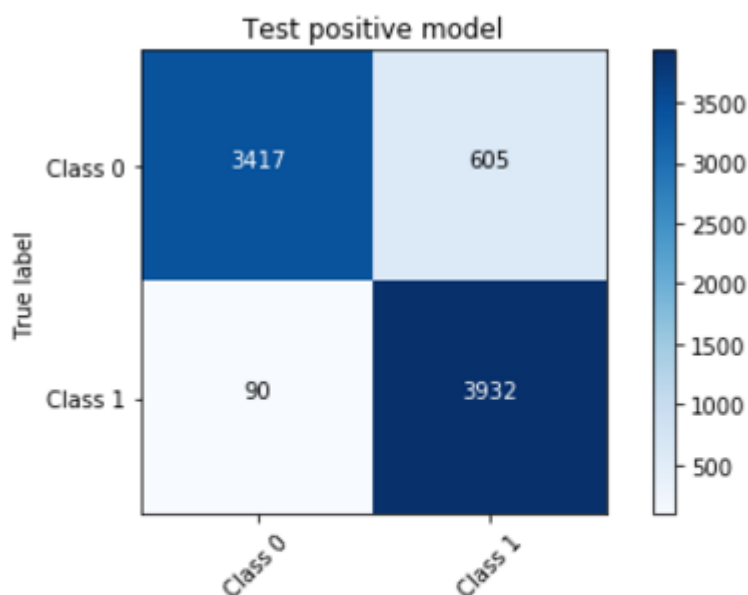
## Результаты обучения модели с позитивной тональностью

Class 0 — сообщения, в которых нет позитивной тональности. Class 1 — сообщения, в которых есть позитивная тональность. Получили достаточно неплохое качество на валидационной выборке из 17062 сообщений — точность 85% для Class 0 и 73% для Class 1.



	precision	recall	f1-score	support
Class 0	0.85	0.67	0.75	8531
Class 1	0.73	0.89	0.80	8531
avg / total	0.79	0.78	0.77	17062

Далее мы попробовали улучшить качество за счет удаления отзывов с отметкой 4. В результате точность повысилась до 97% для Class 0 и до 87% для Class 1.



	precision	recall	f1-score	support
Class 0	0.97	0.85	0.91	4022
Class 1	0.87	0.98	0.92	4022
avg / total	0.92	0.91	0.91	8044

### Проверка качества модели на наших данных

Следующим этапом стала проверка качества модели на наших данных.

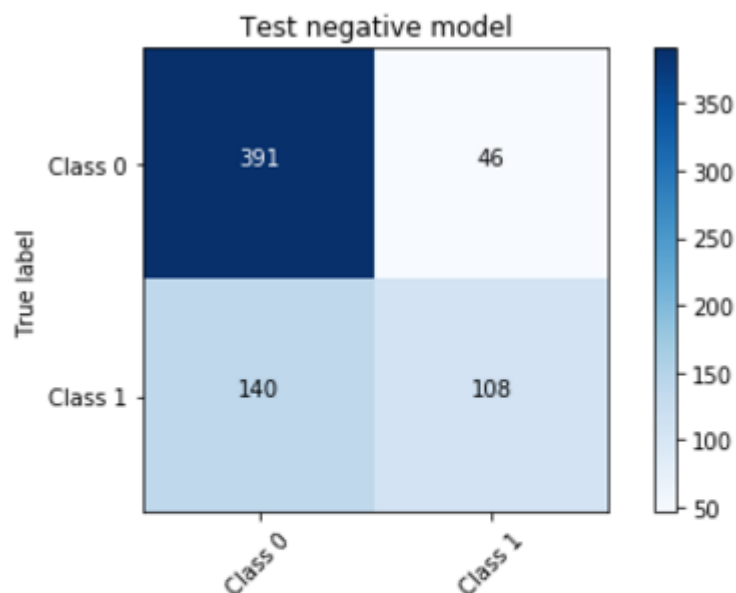
Из проверяемых отзывов были удалены слишком короткие, так как отзывы в обучающей выборке были достаточно объемными и ожидать от модели адекватной работы с текстами из 5-10 слов неправильно, плюс некоторые из них являются мусором.

Примеры коротких отзывов:

#грот #dman55 #концерт #tesla #5летвэфире#ponominalu
Похож на домашний
Понравился
В воскресенье 05.08 в 22.00 отправляю заявку

### Анализ качества негативной модели

Class 0 — сообщения, в которых нет негативной тональности. Class 1 — сообщения, в которых есть негативная тональность. Средняя точность по двум классам — 72%. Но полнота для Class 1 неудовлетворительная — 44%.



	precision	recall	f1-score	support
Class 0	0.74	0.89	0.81	437
Class 1	0.70	0.44	0.54	248
avg / total	0.72	0.73	0.71	685

### Примеры отзывов, на которых модель ошибается

Ниже приведены примеры ложноотрицательных и ложноположительных отзывов, в которых модель допустила ошибки. Некоторые ошибки оправданы, так как есть проблемы в ручной разметке.

#### Ложноотрицательные

При первичной разметке эти отзывы были негативными (Class 1), но модель посчитала их нейтральными (Class 0).

В Руссфине девочка получила 30 тысяч, а в Сетелеме что то не очень зп. Но она не официально, по контракту работает.

Мне нужен был хороший маргарин для выпечки. На полке на глаза попался Спред ЭкоМилк ТМ. Спред изготовлен в Крыму. Состав вполне не плох. Цена за пачку 25 рублей. Запах и сливочный вкус присутствуют. Конечно, в идеале неплохо было бы взять сливочное масло, но этот спред меня вполне устроил. Даже попробовала его с кусочком черного хлеба. Среди спредов, а это обычный маргарин, ЭкоМилк имеет приятный молочный вкус, хорошо намазывается. И вообще, как эконом вариант, может иногда заменить масло. Хотя масло есть масло, а спред — это суррогат, даже с приятным ароматом и вкусом. Каждый сам решает какие продукты будут у него в холодильнике.

Насчет самого магазина я бы не согласился. Очень много косяков в работе. (персонал очень некомпетентен в своей работе, не знает где что лежит, подсказать ничего не могут. Хамство, просрочка, долгое отсутствие товара). Впрочем почти как у всех, но чуть хуже. Подкупает только наличие собственных марок товаров для дома, бизнеса и сада.

#### Ложноположительные

При первичной разметке эти отзывы были нейтральными (Class 0), но модель посчитала их негативными (Class 1).

Сапоги внешне симпатичные, но на полную ногу не подошли. Совершенно не застегнулись в районе щиколотки. Примерила еще и дочь, для узкой ноги вообще везде велики. Отказ.

Ездим сюда регулярно на протяжении нескольких лет. Почему? Ассортимент — в нашем районе ни в одном магазине близко нет такого ассортимента продуктов. Да, цены высокие. Но за выбор готовы платить. Однако в последнее время возможность выбора заметно снизилась, а условия его предоставления выросли. 1. Персонал — при обращении не угадаешь помогут или нахамят (потом возможно помогут), 50\50. В редких случаях могут нахамить без обращения. 2. Ассортимент — периодически целые группы товара отсутствуют на витрине — пустые полки, либо находятся на последнем дне срока эксплуатации (можно бы купить, но об этом дальше) 3. Регулярно попадают товары в пределах срока годности, при распаковке которых на столе оказывается тухлятина. Обращались — было предложено привозить назад. Т.е. покупая товар в этом магазине 1 раз в неделю потребитель должен хранить документы о покупке до вскрытия последнего продукта, после вскрытия хранить испорченный продукт до следующей поездки (сам факт перевозки и обмена тухлятины не воодушевляет), либо срывать в магазин по факту обнаружения — все равно завтрак\обед\ужин испорчен и заняться больше нечем :) Я метровскую тухлятину выкидываю — списываю туда же, к высоким ценам — плата за возможность выбора... Эта картина продолжается на протяжении нескольких лет, что же вынудило написать? Дело в том, что ребята совсем оборзели, открыто выкладывая протухший товар на витрины магазина (хотя возможно это обычное дело, а я просто впервые столкнулся). Фото прилагаю. Самое печальное, что пересматривать свою продуктовую корзину совсем не хочется, т.е. посещать эту «заведение» наверное придется, но возможно данное замечание сможет повлиять на улучшение качества обслуживания в лучшую сторону.

нет не единственный, если мы пришли и потратили купон с 50% скидкой и вместо 470 000 оплатили 235 000,то с 235 000 мне тоже на платиус ничего не начисляется, почему????

#### **Анализ качества позитивной модели**

Class 0 — сообщения, в которых нет позитивной тональности. Class 1 — сообщения, в которых есть позитивная тональность. Средняя точность — 63%. Полнота для Class 0 — 44%, что является неудовлетворительным результатом.



	precision	recall	f1-score	support
Class 0	0.60	0.44	0.51	299
Class 1	0.64	0.77	0.70	386
avg / total	0.63	0.63	0.62	685

### Примеры отзывов, на которых модель ошибается

Ниже приведены примеры ложноотрицательных и ложноположительных отзывов, в которых модель допустила ошибки.

#### Ложноотрицательные

На первичной разметке эти отзывы были положительными (Class 1), но модель посчитала их нейтральными (Class 0).

Начитавшись восторженных отзывов в интернете об этих чудо БАДах, решила и себе приобрести. Покупке не могла нарадоваться, очень давно их хотела. У меня есть некоторые проблемы с ногтями, они ломике, слоятся, медленно растут. Начала пить их по две в день, как положено. Пила месяц, но увы результата нет (немного укрепились ногти, но в целом картина та-же. Очень жаль, но своих денег они не стоят, ибо обычный кальций, который стоит копейки, справляется лучше. Да и пить их нужно строго после еды, потому что тошнота гарантирована.

Ой а у меня сын всё просит пони ему купить. Не такую точно, но тоже радугу эту. Как то это для девочек же, но вот хочет он пони эту и всё.

Выглядят симпатично, но очень жесткие. Вообще не сгибаются. Никогда не видела, чтобы натуральная кожа была такой «деревянной». Отказ. Размер соответствует.

Достоинства: Удобный сайт, легко заказать и оплатить билеты, адекватная наценка, есть самовывоз  
Недостатки: проблемы с возвратом билетов

#### Ложноположительные

На первичной разметке эти отзывы были нейтральными (Class 0), но модель посчитала их положительными (Class 1).

Екатерина, это бонусная карта. Пользуясь системой Platius можно возвращать на свой счет в баллах до 30% стоимости заказа (в зависимости от ранга). А потом этими баллами можно оплачивать до 100% стоимости заказа. Можно активировать при помощи промокода, а можно этой картой.

Девочки, помогайте! По телику идет реклама игрового набора My Little Pony — чемодан красоты или что-то типа того. Вот не помню по какому каналу идет. Может кто запомнил, как точно этот набор называется? Ищу подарки дочке на ДР и НГ

На рынке строительных и отделочных материалов современности наибольшую известность имеют плиточные и рулонные элементы тепловой изоляции. Обычно в категорию плиточных элементов изоляции тепла входят минеральная вата, изолвер, стиродур, пеноплекс, пенопласт, а также пенополистирол. В категорию рулонных элементов изоляции тепла входят вспененный полиэтилен и пенофол.

Видно, что в группе ложноотрицательных отзывов есть примеры, в которых модель сделала правильный выбор, так как явно не указаны достоинства или указано незначительное преимущество. Таких примеров около 10% в ошибочных классификациях, поэтому качество модели несколько выше, чем показывает матрица неточностей.

В целом качество работы модели на наших данных не очень хорошее. Это объясняется несколькими факторами:

1. Для обучения брались отзывы, у которых только одна тональность, а в нашей разметке один отзыв может иметь и положительную и негативную тональности.

2. Отзывы на сайте [otzovik.com](http://otzovik.com) проходят модерацию и имеют определенный формат: средняя длина текста составляет 200 слов, отсутствует ненормативная лексика, оскорбления. Мы же анализируем отзывы из разных источников, в том числе из Twitter, где длина текста сильно ограничена.

### **Выводы**

Модели имеют очень хорошее качество на длинных отзывах, которые проходят модерацию, поэтому для автоматической разметки некоторых источников они могут использоваться прямо сейчас.

Для повышения качества обработки коротких и немодерируемых отзывов требуется доработка моделей: в первую очередь, матрицу TF—IDF нужно составлять с использованием сторонних источников, чтоб модели умели обрабатывать большее количество слов.