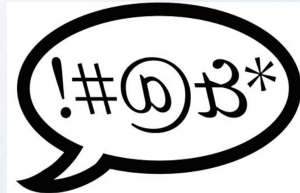


# Анализ и предсказание тональности сообщений



Авторы студенты группы  
ИВТ-22

Богодухов А.

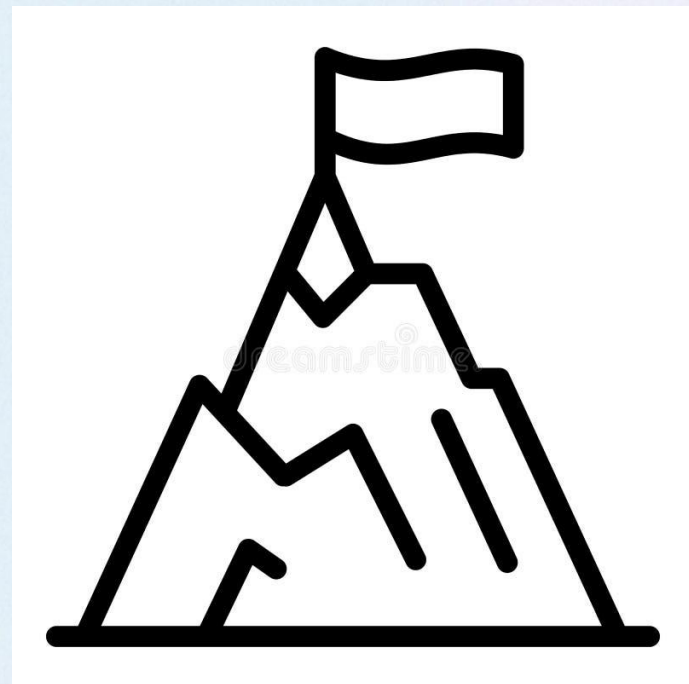
Нечталенко И.

Сапожников Ю.

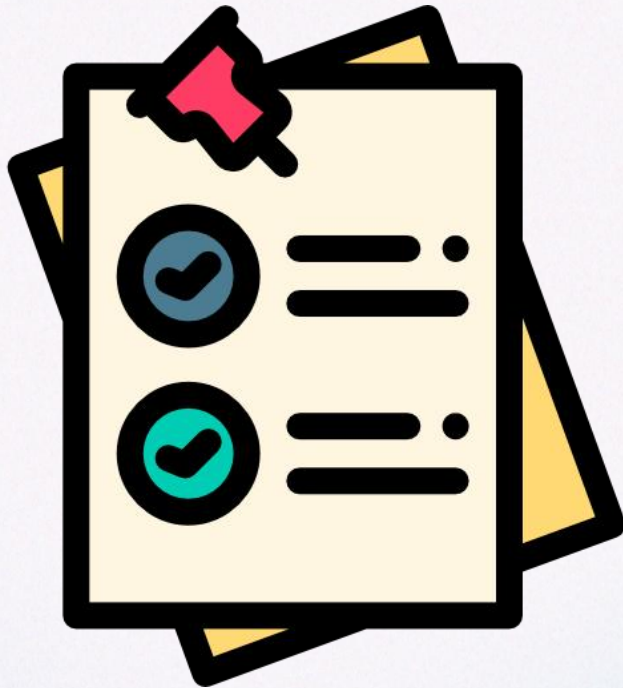
## Цель проекта

Цель разработать метод предсказания тональности сообщений с использованием методов анализа данных, понять, как люди воспринимают продукты, события или бренды.

В этом могут помочь различные алгоритмы машинного обучения и нейронные сети для решения задачи классификации.



# Актуальность



Анализ тональности используется для понимания настроений пользователей в социальных сетях, отзывах и комментариях.

В различных компаниях эти данные используют для улучшения качества обслуживания клиентов и адаптации своих стратегий маркетинга.



# Описание задачи

Целью является предсказания тональности заключается в автоматической классификации текстовых сообщений по категориям: положительная, отрицательная или нейтральная.

Автоматизация нужна для обработки больших объемов данных быстрее и точнее, чем при ручной классификации.



# Что такое тональность

Тональность текста – это отношение некоторого лица к теме текста или каким-то ее аспектам. Тональность может быть положительной, отрицательной или нейтральной.

**Положительная тональность** - текст передает позитивные эмоции, выражает одобрение или удовлетворение.

**Отрицательная тональность** - текст выражает негативные эмоции, критику или неудовлетворение.

**Нейтральная тональность** - текст не содержит выраженной эмоциональной окраски, чаще всего сообщает факты.

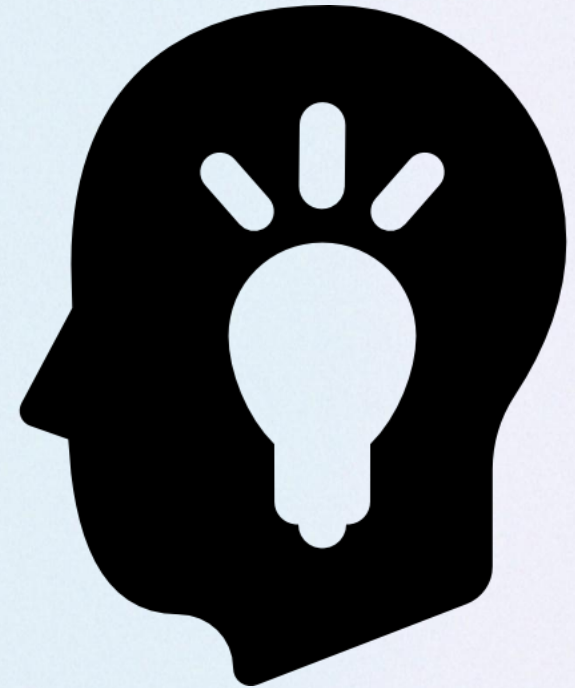
## Как определить тональность сообщений

Для определения тональности сообщения используют несколько методов, такие как:

**Лексический анализ** - использование заранее созданных словарей тональности, где каждому слову присвоена положительная, отрицательная или нейтральная оценка.

**Машинное обучение** - обучение на размеченных данных: Алгоритмы анализируют тысячи примеров с заранее помеченной тональностью и учатся определять её в новых сообщениях. Такие как: модели, такие как логистическая регрессия, SVM, деревья решений.

**Анализ с помощью нейронных сетей** - рекуррентные нейронные сети (RNN) и их улучшенные варианты (LSTM, GRU) анализируют порядок слов и их взаимосвязь в тексте.





# Проблемы определения тональности

## Сложности с контекстом:

Тональность некоторых слов зависит от контекста. Например, слово "хорошо" может иметь саркастический оттенок: "Ну, конечно, очень хорошо, что ты опоздал!" – здесь очевидно отрицательная тональность, несмотря на использование положительного слова.

## Сарказм и ирония:

Алгоритмам сложно правильно интерпретировать саркастические или ироничные высказывания.

Пример: "Да, мне так нравится, когда поезд задерживается на час". Здесь выражается недовольство, но фраза выглядит как положительная.



# Описание датасета

Для анализа данных был выбран датасет AlexSham/Toxic\_Russian\_Comments. Датасет состоит из 248290 строк и разделен на 2 колонки text и label. В столбце text хранятся тексты комментариев, в label представлены метки для каждого, где 0 - нейтральные комментарии, а 1 - токсичные комментарии пользователей.

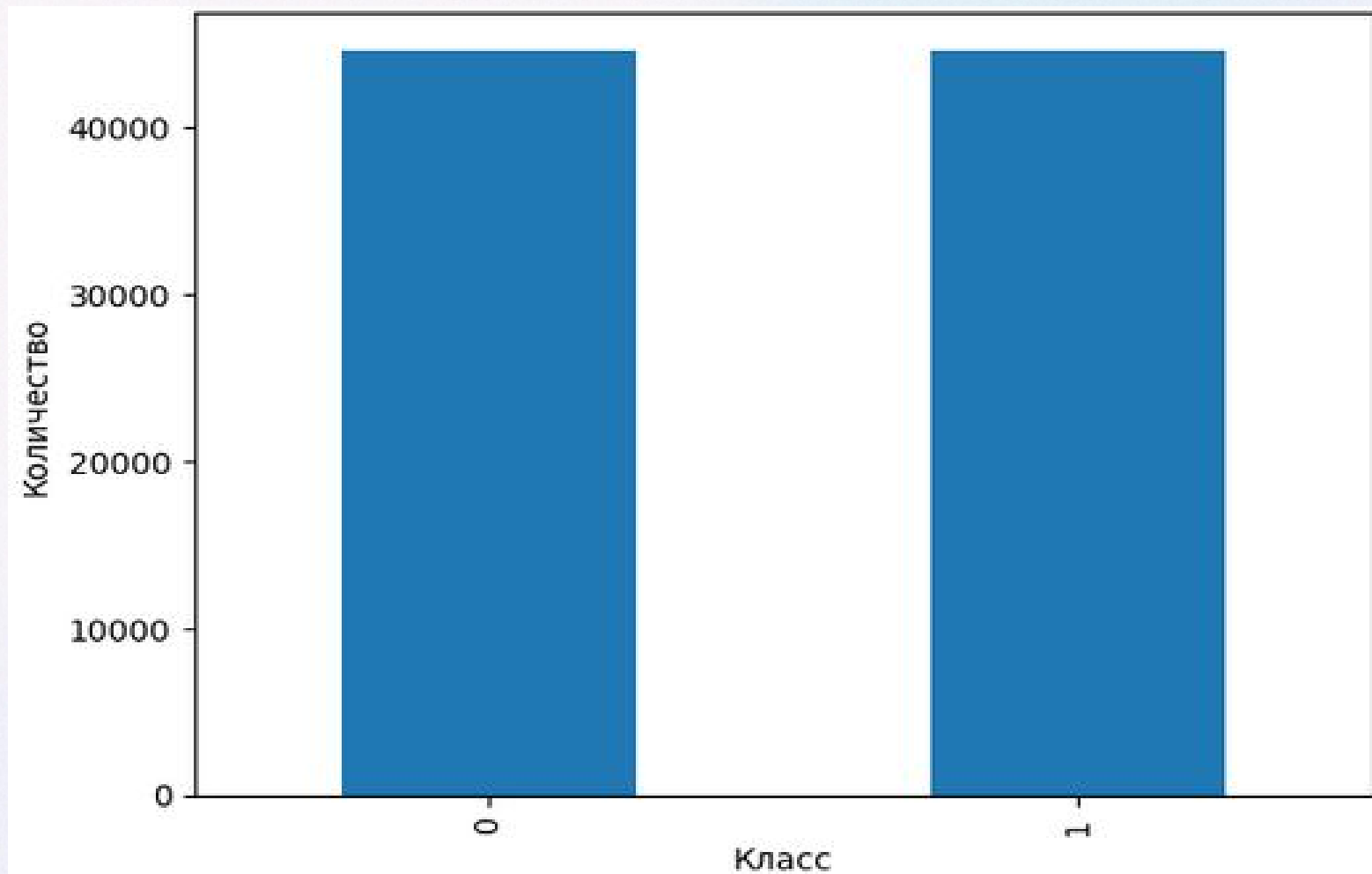
	text	label
0	видимо в разных регионах называют по разному ,...	0
1	понятно что это нарушение правил, писать капсл...	1
2	какие классные, жизненные стихи....	0
3	а и правда-когда его запретили?...	0
4	в соленой воде вирусы живут .ученые изучали со...	0
...	...	...
248285	и ты будь здоров	0
248286	не дорога а прям стекло но правда битое (h)	0
248287	спасибо большое. буду ждать хороших новостей. ...	0
248288	активирую установку 🌈🌈🌈👍😎🔥🔥🔥	0
248289	а вы курс российского рубля видели, кошмар!!!	0

248290 rows × 2 columns

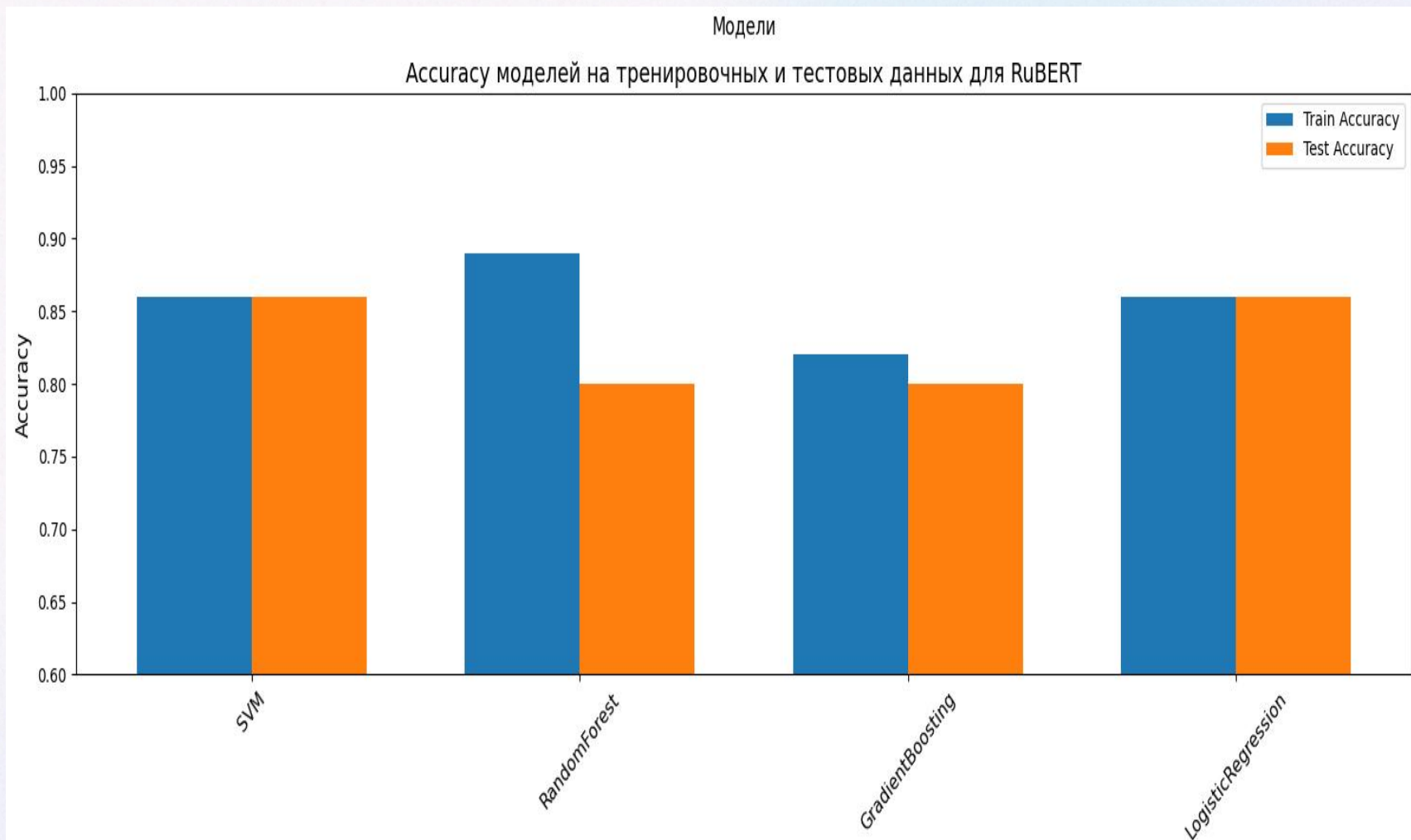


# Предварительная обработка данных

Была проведена очистка данных и также были сбалансированы оба класса



# Оценка качества модели



## Заключение

В заключений можно сказать что анализ тональности сообщений в интернете помогает понять мнения пользователей о брендах и продуктах, выявляя общественное настроение. Несмотря на сложности, такие как сарказм и многозначность, этот инструмент находит широкое применение в маркетинге и управлении репутацией, способствуя принятию обоснованных решений.



4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0

# Спасибо за внимание



**Заблокированный пользователь**

@id250229257

Данный материал заблокирован на  
территории Российской Федерации по  
требованию Роскомнадзора 1057628-НВ