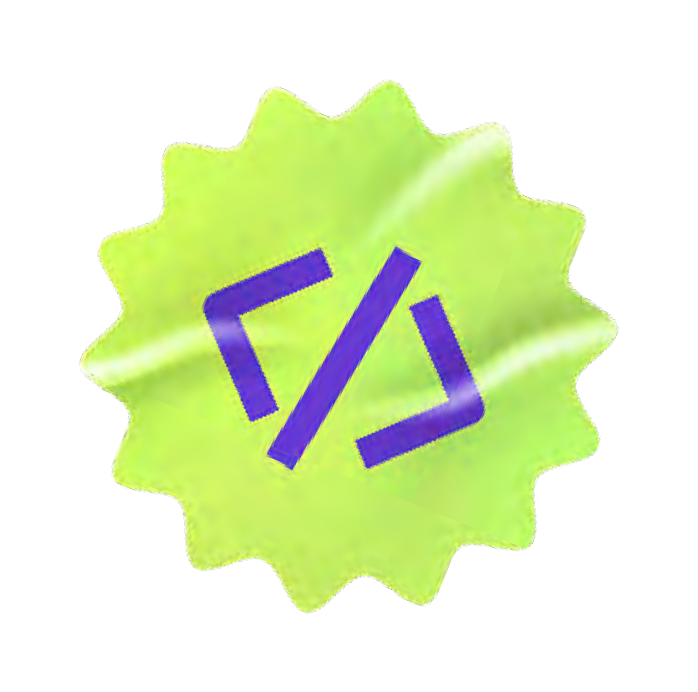


Анализатор тональности отзывов

### Ключевая задача кейс-чемпионата — разработка системы для автоматического анализа тональности русскоязычных текстов, реализованной в виде полнофункционального веб-приложения. Проект направлен на решение практической задачи по классификации пользовательских отзывов и комментариев по трем категориям: отрицательная, нейтральная и положительная эмоциональная окраска.

**Оглавление**



Контекст задачи [3](#_bookmark1)

Постановка задачи [4](#_bookmark2)

[Детализация технических требований к решению 5](#_TOC_250001)

Ограничение пространства решений [6](#_bookmark4)

Заинтересованные стороны [6](#_bookmark4)

Источники информации [7](#_bookmark5)

[Критерии оценки решения на хакатоне](#_TOC_250000) 7

[Об организации 8](#_bookmark6)

[Карьерные возможности 9](#_bookmark7)

## [Приложение 10](#_bookmark8)

Команда Changellenge >> подготовила этот кейс исключительно для использования в образовательных целях.

Авторы не намерены иллюстрировать как эффективное, так и неэффективное решение управленческой проблемы. Кейс не содержит исчерпывающую информацию, необходимую для решения. Вы можете использовать любые дополнительные источники и свои допущения.

Некоторые имена в кейсе, а также другая идентификационная информация могли быть изменены с целью соблюдения конфиденциальности. Changellenge >> запрещает любое неправомерное воспроизведение, хранение или распространение кейса без письменного разрешения. Для заказа копии, получения разрешения на использование или в случае обнаружения неправомерного использования данного кейса, пожалуйста, свяжитесь с нами по адресу [info@changellenge.com.](mailto:info@changellenge.com)

[Анализатор тональности отзывов](#_bookmark0) 2

# Контекст задачи



##### Отзывы горожан — ценная информация, которая помогает столице развивать городские цифровые сервисы. Но анали- зировать вручную большой объем комментариев — трудо- емкая и практически невыполнимая задача для масштабов такого мегаполиса, как Москва. В связи с этим актуальной становится разработка и внедрение автоматизированных методов анализа тональности текста.

##### В рамках данной задачи требуется разработать модель ма- шинного обучения, способную классифицировать тональ- ность русскоязычных текстов, а также создать веб-прило- жение с интерфейсом для взаимодействия с моделью. Такой сервис поможет городу быстрее классифицировать обраще- ния и оперативнее узнавать, что в цифровых проектах столи- цы нравится горожанам, а что требует улучшения.

[Анализатор тональности отзывов](#_bookmark0) 3

# Постановка задачи



#### Разработайте модель классификации текстов по эмоциональной тональности (отрицательная, нейтральная, положительная) и представьте результат в форма- те web-интерфейса.

* 1. Базовая функциональность, которой должен обладать итоговый продукт:
     + Разработанный интерфейс позволяет загружать данные и проводить автоматическую нормализа- цию текста, например токенизацию, лемматизацию, NER и др.
     + Модель распознает эмоциональную тональность текста и возвращает оценку от 0 до 2, где 0 — отрицательная, 1 — нейтральная, 2 — положительная.
     + Интерфейс возвращает размеченные данные в формате csv-файла и визуализацию итогового результата.
     + Интерфейс предоставляет возможность оценить размеченные моделью данные по метрике macro-F1 путем добавления csv-файла с валидационной выборкой (она эквивалентна тестовой выборке, но с целевой колонкой, размеченной экспертом)
  2. Дополнительные фичи:
     + Поиск текстов по их источникам, а также конкретным словоформам в текстах.
     + Фильтрация текстов по источникам, оценкам.
     + Ручная корректировка разметки.
     + Возможность выбора различных визуализаций данных.

[Анализатор тональности отзывов](#_bookmark0) 4

# Детализация технических требований к решению

1. Чего хотим

Итоговое решение должно представлять собой работающий прототип веб-приложения, позволя- ющий провести классификацию входных данных в зависимости от их эмоциональной тонально- сти и создать визуализации итогового результа- та. Дополнительным плюсом станет возможность фильтрации и поиска по текстам, ручной кор- ректировки разметки модели классификатора, а также широкий выбор визуализаций, например дашборд с классификацией отзывов или динами- кой изменения тональности.

1. Как это должно работать

С точки зрения backend обязательно создание модели для задачи классификации текстов на русском языке по критерию эмоциональной то- нальности (негативный, нейтральный, положи- тельный), автоматическая нормализация входных данных, компиляция итогового файла размечен- ных данных в формате CSV.

С точки зрения frontend мы ожидаем получить веб-интерфейс, который можно использовать для подгрузки входных данных в формате CSV, про- смотра итогового результата и визуализаций, вы- грузки файла CSV с итоговой разметкой текстов, проведения оценки итогового результата размет- ки по метрике macro-F1 путем добавления вали- дационной выборки данных в формате CSV.

1. Что по ограничениям

Алгоритм классификации должен быть реализо- ван таким образом, чтобы его выполнение завер- шалось за разумное время при использовании небольших вычислительных ресурсов.

1. На что обращаем внимание

Обеспечьте кроссплатформенную совместимость веб-приложения. Протестируйте его работу на различных устройствах и основных операцион- ных системах. Убедитесь в надежности взаимо- действия между frontend и backend, включая кор- ректность обмена данными и обработку ошибок.

Постройте модель с использованием библиотек и фреймворков и обучите модель на предостав- ленных данных. Модель должна работать на обычном железе (стандартные cpu/gpu).

1. В каком виде предоставить результат

Готовый продукт должен включать:

**Репозиторий.** Разместите исходный код бэкэнда, фронтенда, а также код обучения модели в от- крытом репозитории Mos.Hub, создав публичный проект по ссылке [hub.mos.ru](https://hub.mos.ru/users/sign_in).

**Инструкцию по локальному развертыванию.** Представьте руководство по запуску проекта, включающее загрузку файлов, настройку рабо- чего окружения и развертывание необходимой среды для полноценного локального тестирова- ния проекта.

**Размещение на бесплатном хостинге.** Заде- плойте ваше готовое веб-приложение на одном из бесплатных облачных сервисов и предоставьте ссылку на ваш развернутый проект.

**CSV-файл** размеченного тестового датасета и от- чет о разметке по метрике macro-F1.

**Контейнеризацию (опционально).** Дополни- тельно ко всему вышеописанному вы можете предоставить доступ к вашему продукту, собран- ному с использованием инструментов контейне- ризации, таких как Docker.

[Анализатор тональности отзывов](#_bookmark0) 5



# Ограничение пространства решений

Стек технологий: Python для построения модели (библиотеки Pandas, Numpy, Matplotlib, PyTorch, TensorFlow) и пр.

Использование открытых моделей допускается

и приветствуется.

Применение закрытых моделей или проприетарных API запрещено.

Использование малых языковых моделей (в том числе мультиязычных, адаптированных под русский язык) допускается.

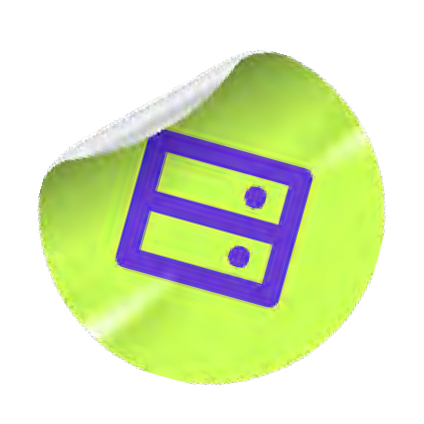
# Заинтересованные стороны

Жители,

чье мнение, выраженное в отзывах и комментариях, будет услышано и проанализировано, что в перспективе приведет к улучшению качества товаров и услуг.

Разработчики IT-проектов,

заинтересованные в получении обратной связи от пользователей.



[Анализатор тональности отзывов](#_bookmark0) 6

# Источники информации

**1.**

[Анализ тональности текста: зачем он нужен и как его](https://habr.com/ru/companies/data_light/articles/851700/) [использовать?](https://habr.com/ru/companies/data_light/articles/851700/)

**2.**

Анализ тональности в русскоязычных текстах (части [1](https://habr.com/ru/companies/vk/articles/516214/), [2](https://habr.com/ru/companies/vk/articles/516726/) и [3](https://habr.com/ru/companies/vk/articles/516730/)).

# Критерии оценки решения на хакатоне

Результаты оцениваются по нескольким критериям:

1. Точность модели классификатора по macro-F1. Чем выше данная метрика, тем лучше. Метрика — macro-F1.

Формула (по классам):

Пример вычисления в Python:

from sklearn.metrics import f1\_score

Σ

2 • Precisioni • Recalli

1 N score = f1\_score(y\_true, y\_pred, average='macro'

F1i=

Precisioni

+ Recalli

Macro – F1 =

N

F1i

i = 1



1. Запускаемость проекта. **3.** Frontend: user-friendliness, качество и наглядность визуализаций.

[Анализатор тональности отзывов](#_bookmark0) 7

# Об организации



##### Департамент информационных технологий Москвы (ДИТ) вместе со всеми городскими ведомствами создает и разви- вает цифровые проекты, которые делают Москву комфортнее, а жизнь горожан — удобнее и мобильнее.

##### Для системы управления столицей технологии — это незаме- нимый инструмент, применяемый во всех отраслях экономи- ки, городского хозяйства и социальной сферы. А для миллио- нов горожан — повседневный помощник, который позволяет получать сотни услуг и сервисов в удобном цифровом фор- мате в режиме 24/7/365.

[Анализатор тональности отзывов](#_bookmark0) 8

# Карьерные возможности

##### До 12 декабря идет прием заявок на стажировку «IT-город». https://moscowcareer.mguu.ru/itgorod/

##### Стажировка «IT-город» в Правительстве Москвы — это шанс попасть в команду, которая с помощью современных техно- логий делает столицу еще комфортнее для жизни.

Старт программы — 2 марта 2026 года. Длительность — 6 месяцев.

##### Стажировка проходит в департаменте информационных технологий Москвы по десяти направлениям:

Аналитика данных Информационная безопасность

Машинное обучение Разработка Тестирование

Управление проектами Бизнес и системный анализ

Дизайн Системное администрирование

Поддержка пользователей и систем

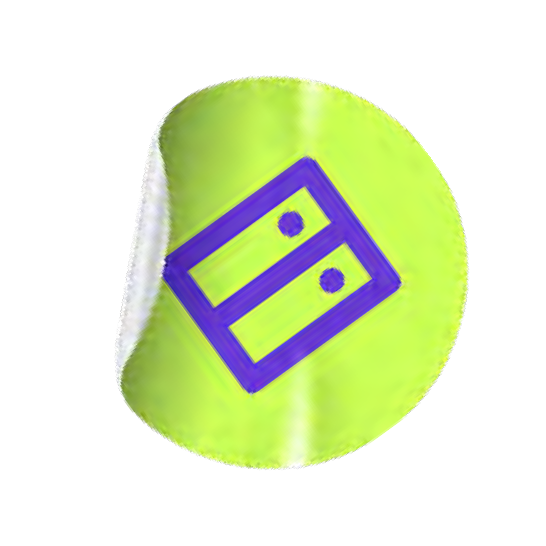
##### Что ждет на стажировке:

Участие в развитии столичных цифровых проектов: портала mos.ru, ЕМИАС, МЭШ и многих других

Зарплата до 75 000 рублей в зависимости от выбранного направления и количества отработанных часов

Гибкий график: можно совмещать учебу и стажировку

Карьерные перспективы: более 60% стажеров остаются работать в IT-команде столицы



[Анализатор тональности отзывов](#_bookmark0) 9

**Приложение**

Архив с датасетом включает в себя:

* train.csv — обучающий набор (текст отзыва + тональность + источник текста);
* test.csv — тестовый набор (без целевой колонки);
* sample\_submission.csv — пример формата сдачи итогового CSV-файла.

##### [Ссылка](https://drive.google.com/file/d/1e-sTvQRZsx2dPjAvkjRSdbWDQ1ra1zAJ/view?usp=sharing)

Структура train.csv, где:

* text — текст отзыва;
* label — числовая метка тональности (0, 1, 2);
* src — источник данных (например, rureviews).

text

Пальто красивое, но пришло с дырой в молнии

label 0

src rureviews

Очень долго шел заказ, ждала к Новому году, 0

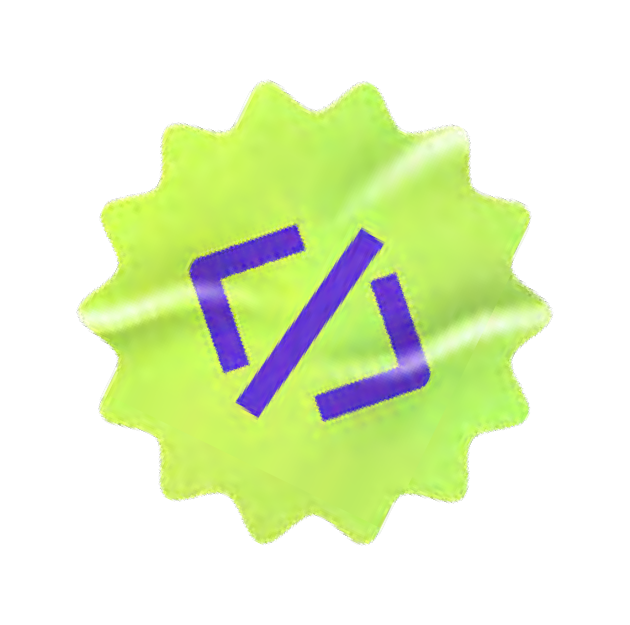
пришел после праздников

rureviews

Доставка быстрая, меньше месяца 1

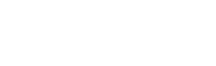
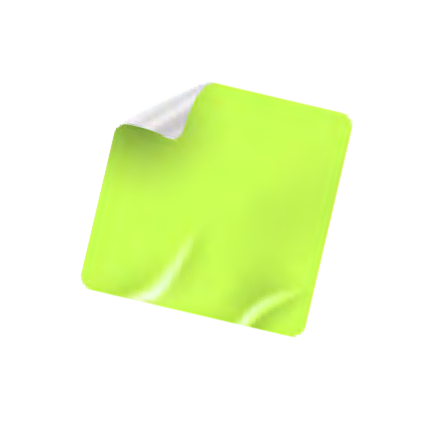
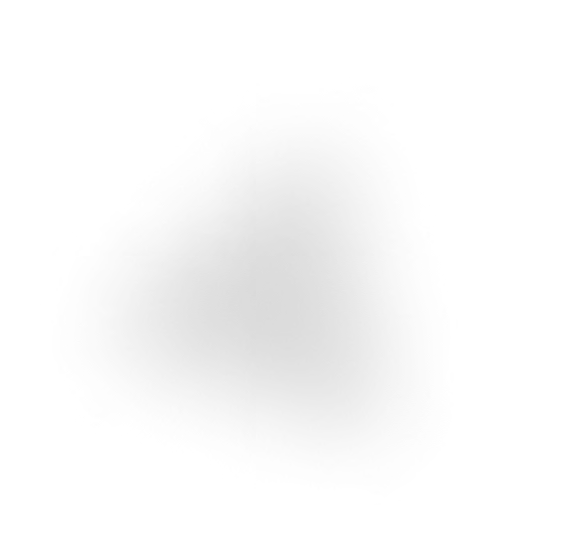
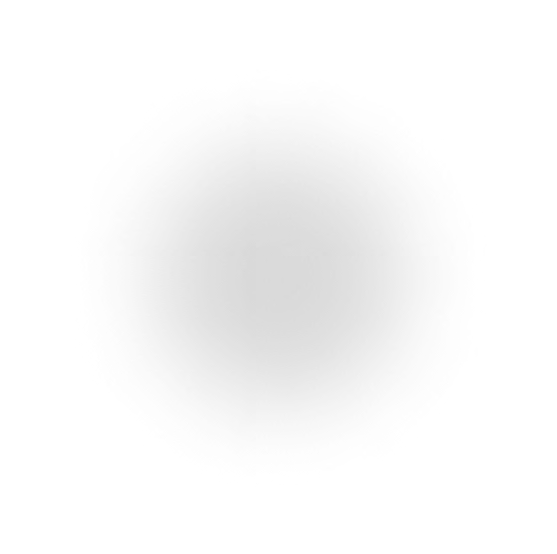
Все отлично, качество супер! 2

rureviews rureviews



[Анализатор тональности отзывов](#_bookmark0) 10

###### Кейс написан и опубликован Changellenge >> — ведущей организацией по кейсам в России.



###### [www.changellenge.com](http://www.changellenge.com/) [info@changellenge.com](mailto:info@changellenge.com) vk.com/changellengeglobal

###### Кейс написан по заказу Правительства Мосвы

###### [www.mos.ru](http://www.mos.ru/)