

**цифровой  
прорыв** 

**сезон: ИИ**

# **ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗАТОР ОБРАТНОЙ СВЯЗИ СТУДЕНТОВ**

**Команда: `random.randname()`**

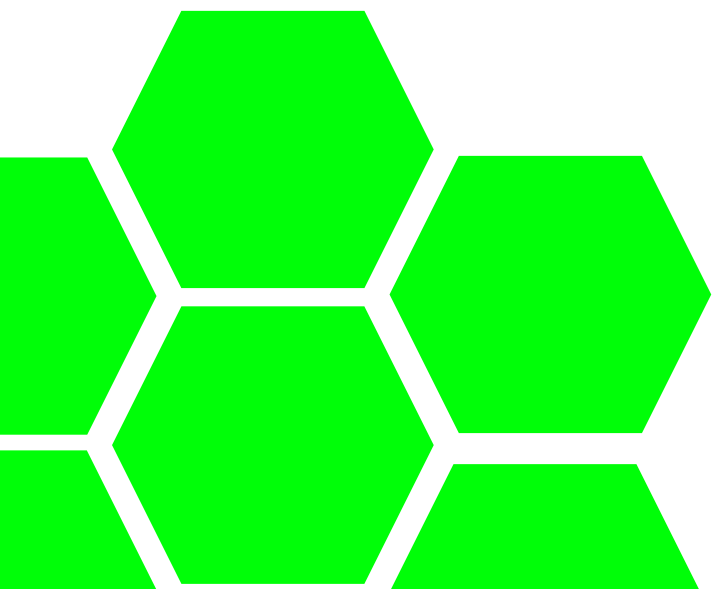
**Аминев Роман**

**Белоусов Максим**

**Левитская Диана**

**Серов Иван**

**Строкова Анастасия**



## СОДЕРЖАНИЕ

Постановка задачи .....	3
Схема базы данных .....	4
ML-решение .....	7
Telegram-бот .....	8
Модуль отчетов .....	9
UML-диаграммы .....	12

## ПОСТАНОВКА ЗАДАЧИ

### Условие задачи

Для повышения эффективности обучения и улучшения качества программ требуется обрабатывать большой объем обратной связи. Для этого необходимо разработать прототип системы для интеллектуального анализа обратной связи студентов. Система должна включать в себя чат-бота для сбора данных от студентов и алгоритмы искусственного интеллекта для обработки и классификации полученной информации. Чат-бот задает студентам открытые вопросы после вебинаров и сессий, собирая ответы в текстовом формате. Далее алгоритмы ИИ анализируют собранные ответы, выделяя ключевые моменты и классифицируя их на информативные и неинформативные, а также на положительные и отрицательные отзывы.

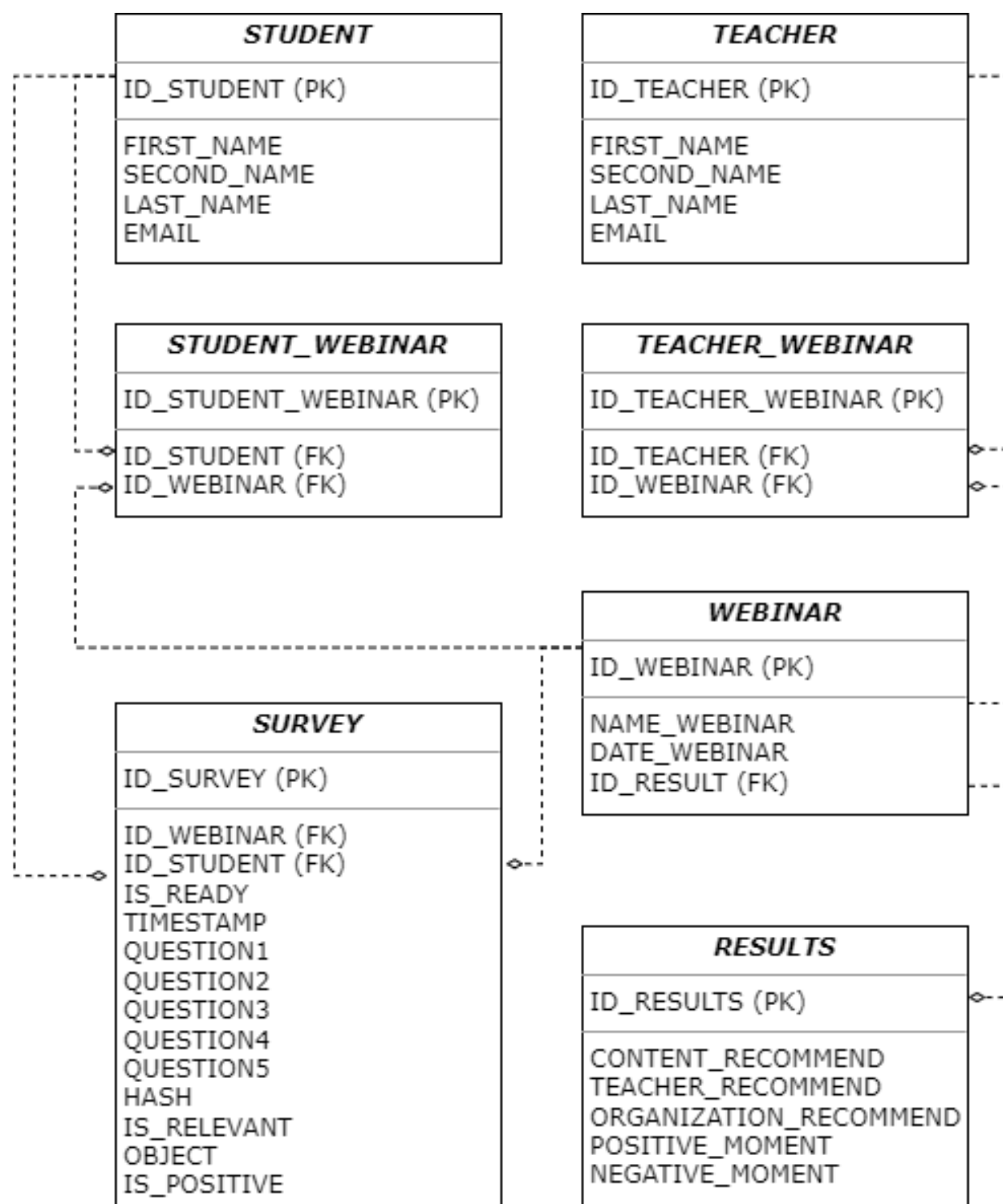
### Общая идея решения

Чат-бот OrnioMate, предназначенный для анализа преподавателем или автором курса обратной связи студентов по вебинарам. Бот использует технологии искусственного интеллекта для классификации ответов студентов по нескольким категориям для более удобной обработки и извлечения полезной информации. Главной технической особенностью бота является качество его классификации и больше количество визуальных интерпретаций фидбека, что делает его идеальным инструментом для выявления точек роста образовательных курсов. Бот предоставляет уникальную методику создания визуальных и текстовых отчетов, подготовленных для анализа.

### Преимущества решения

- *Ролевая модель* (преподаватель + студент): модель позволяет использовать программное обеспечение как студентам, предоставляющих ответы на открытые вопросы, так и преподавателям, получающим рекомендации по проведению вебинаров.
- *Внесение информации о новых вебинарах*: пользователь с ролью «Преподаватель» может самостоятельно через Telegram-бот занести информацию о наименовании и дате планируемого вебинара. Это позволяет повысить гибкость управления программным обеспечением (no-code).
- *Точность и эффективность моделей*: в процессе разработки были проведены несколько экспериментов с нейронными сетями, позволившие получить высокие метрики качества (в том числе благодаря обогащению датасета).
- *Легкость интерпретации данных*: разнообразие графиков позволяют визуально понять сильные стороны, а также точки роста.
- *Использование широко распространенного мессенджера для работы*: Telegram на сегодняшний день в России обгоняет по популярности Whats App и Viber, что говорит о легкости использования бота (нет потребности в дополнительной установке программ на мобильное устройство).

## СХЕМА БАЗЫ ДАННЫХ



**STUDENT**

Студент

Атрибут	Тип данных	Комментарий
ID_STUDENT (PK)	Integer	Уникальный идентификатор
FIRST_NAME	String	Имя
SECOND_NAME	String	Фамилия
LAST_NAME	String	Отчество
EMAIL	String	Электронная почта

**TEACHER**

Преподаватель

Атрибут	Тип данных	Комментарий
ID_TEACHER (PK)	Integer	Уникальный идентификатор
FIRST_NAME	String	Имя
SECOND_NAME	String	Фамилия
LAST_NAME	String	Отчество
EMAIL	String	Электронная почта

**WEBINAR**

Вебинар

Атрибут	Тип данных	Комментарий
ID_WEBINAR (PK)	Integer	Уникальный идентификатор
NAME_WEBINAR	String	Наименование
DATE_WEBINAR	Date Time	Дата
ID_RESULT (FK)	String	Связь с таблицей с рекомендациями

**STUDENT\_WEBINAR**

Устранение связи многие-ко-многим (студент может принимать участие во многих вебинарах, на вебинаре могут присутствовать несколько студентов)

Атрибут	Тип данных	Комментарий
ID_STUDENT_WEBINAR (PK)	Integer	Уникальный идентификатор
ID_STUDENT (FK)	Integer	Уникальный идентификатор
ID_WEBINAR (FK)	Integer	Уникальный идентификатор

**TEACHER\_WEBINAR**

Устранение связи многие-ко-многим (один преподаватель может вести много вебинаров, в одном и том же вебинаре могут обучать несколько преподавателей-спикеров)

Атрибут	Тип данных	Комментарий
ID_TEACHER_WEBINAR (PK)	Integer	Уникальный идентификатор
ID_TEACHER (FK)	Integer	Уникальный идентификатор
ID_WEBINAR (FK)	Integer	Уникальный идентификатор

## SURVEY

Анкета студента и столбцы, которые мы будем выводить в submission

Атрибут	Тип данных	Комментарий
ID_SURVEY (PK)	Integer	Уникальный идентификатор
ID_WEBINAR (FK)	Integer	Уникальный идентификатор
ID_STUDENT (FK)	Integer	Уникальный идентификатор
IS_READY	Integer	Балл от 1 до 10 о готовности к прохождению опроса
TIMESTAMP	Date Time	Время из train
QUESTION1	String	Первый вопрос
QUESTION2	String	Второй вопрос
QUESTION3	String	Третий вопрос
QUESTION4	String	Четвертый вопрос
QUESTION5	String	Пятый вопрос
HASH	Integer	Хэш
IS_RELEVANT	Boolean	Релевантность ответа (1 – да, 0 – нет)
OBJECT	Integer	
IS_POSITIVE	Boolean	Окраска отзыва (1 – положительный, 0 – отрицательный)

## RESULTS

Обработанные результаты анкеты (текстовые рекомендации о вебинаре, возможно потребуется корректировка БД в этой части после выборов алгоритмов ИИ)

Атрибут	Тип данных	Комментарий
ID_RESULTS (PK)	Integer	Уникальный идентификатор
CONTENT_RECOMMEND	String	Рекомендации о наполненности курса, контенте
TEACHER_RECOMMEND	String	Рекомендации о преподавателе
ORGANIZATION_RECOMMEND	String	Рекомендации об организации
POSITIVE_MOMENT	String	Общие положительные моменты
NEGATIVE_MOMENT	String	Общие негативные моменты

## ML-РЕШЕНИЕ

**Классификация отзывов на информативные/неинформативные и позитивные/негативные**


Инструментарий:

- *RuBERT* – адаптация популярной языковой модели BERT для русского языка, использована для классификации текста (определение тональности, релевантности, тематики);
- *AutoTokenizer* – предобработка текста, обеспечение согласованности и совместимости между токенизацией и предварительно обученной моделью;
- *AutoModelForSequenceClassification* – решение задачи определения категории или метки для входной последовательности текста.

Разработаны следующие функции:

- *classify* – классифицирует с помощью переданного классификатора (path: str - путь к модели, return: int - предсказанная метка)
- *classify\_relevant* - классифицирует по информативности (answers: list[str] - ответы студента на вопросы, return: int - метка класса по релевантности)
- *classify\_positive* - классифицирует по сентименту (answers: list[str] - ответы студента на вопросы, return: int - метка класса по сентименту)
- *classify\_object* - классифицирует по объекту отзыва (answers: list[str] - ответы студента на вопросы, return: int - метка класса по объекту отзыва)

Демонстрация работы модели:

	Reviews	is_relevant	is_positive	object
5714	доступно и понятно все	0	1	1
5715	шикарнейший курс, во многом помог подготовиться...	1	1	1
5716	Курс хороший, но есть местами недочеты.	0	0	1
5717	хороший курс	0	1	1
5718		0	1	1

## TELEGRAM-БОТ

## Telegram-бот

Инструментарий:

- *pyTelegramBotAPI* – создание ботов Telegram с различными функциями, такими как отправка и получение сообщений, обработка команд, inline-клавиатур, файлов и др.
- *AudioSegment* – работа со звуковыми данными для распознавания голосовых сообщений, если студент предоставляет обратную связь с помощью этой функции.
- *mysql.connector* – работа с базами данных MySQL в Python.

Классы:

- *User* – хранение информации о пользователе (имя, фамилия, email, роль)
  - *process\_firstname\_step* – получает и сохраняет имя пользователя
  - *process\_secondname\_step* – получает и сохраняет фамилию пользователя
  - *process\_lastname\_step* – получает и сохраняет отчество пользователя
  - *process\_email\_step* – получает и сохраняет email пользователя
  - *process\_role\_step* – получает и сохраняет роль пользователя (студент или преподаватель)
- *Webinar* – хранение информации о вебинаре (название, дата)
  - *add\_webinar\_info* – проверяет, является ли пользователь преподавателем
  - *process\_webinar\_step* – создание новых вебинаров.

Демонстрация работы:

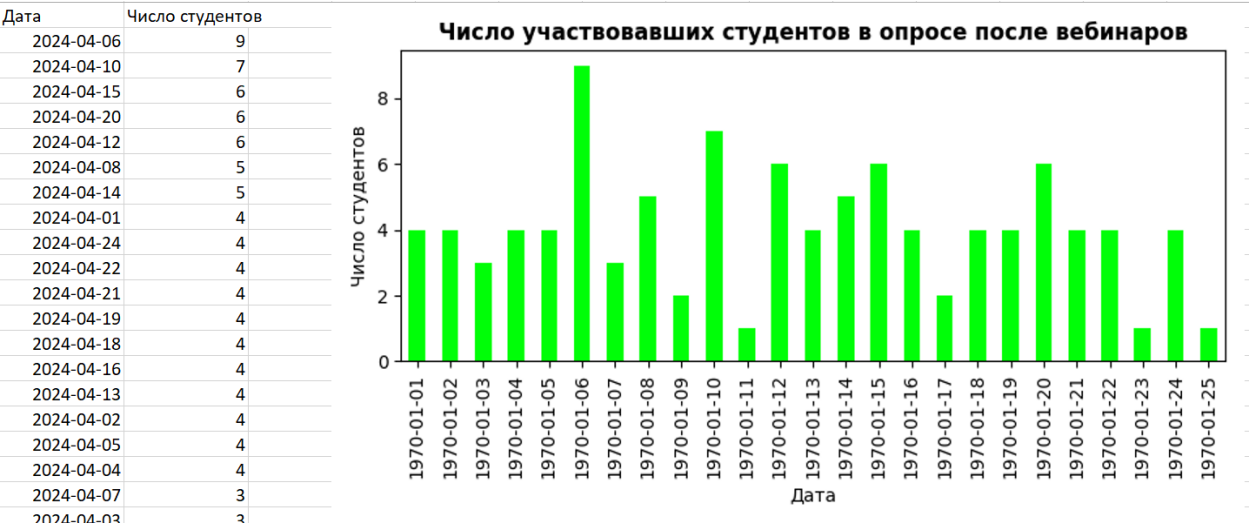




МОДУЛЬ ОТЧЕТОВ

Разработанная визуализация

- График числа участвовавших студентов в опросе после конкретного вебинара  
*survey\_statistics*



- Круговая диаграмма по соотношению релевантности отзывов к вебинарам  
*relevance\_of\_reviews*



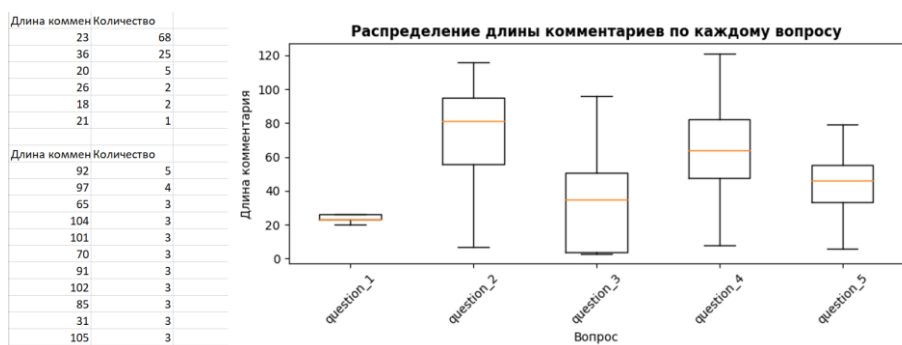
- Круговая диаграмма по соотношению положительных/отрицательных отзывов

*positive\_or\_negative*



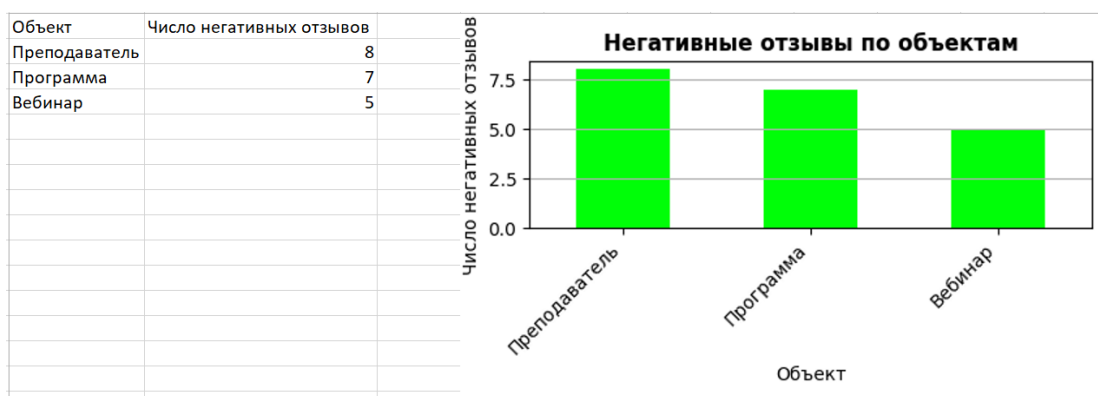
- Распределение длины комментариев по каждому вопросу

*distribution\_of\_response\_lengths*



- Объекты негативных моментов

*negative\_reviews\_of\_the\_objects*

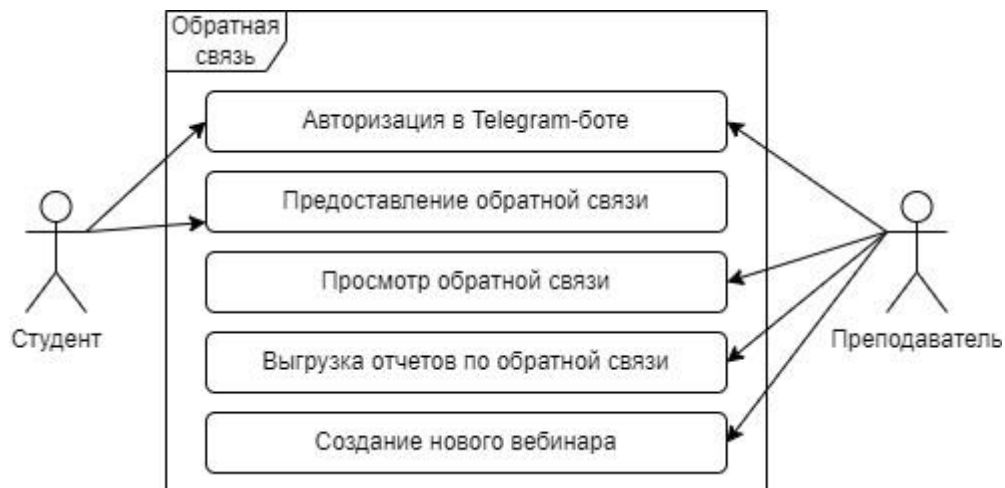


Также были выведены облака слов и биграммы по каждому из вопросов. Пример тем, которые хотят дополнительно изучить студенты (question\_5)



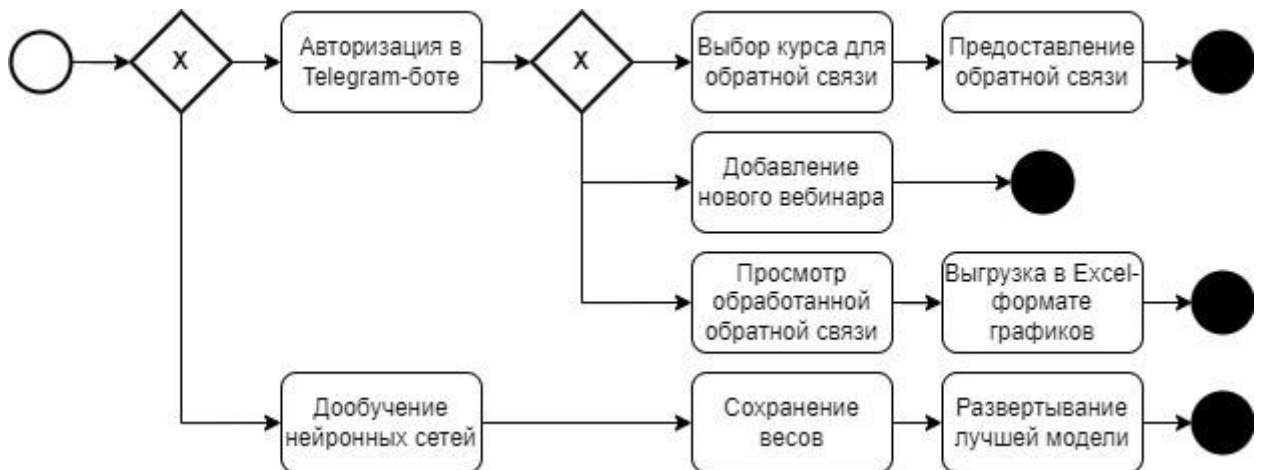
## UML-ДИАГРАММЫ

## Диаграмма прецедентов

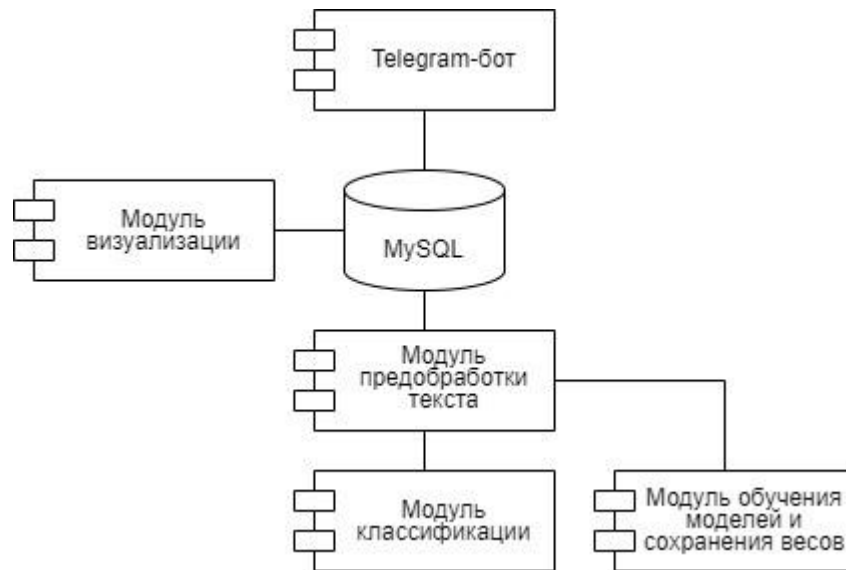


Во взаимодействии с Telegram-ботом участвуют два актора: Студент (авторизуется и предоставляет обратную связь) и Преподаватель (авторизуется, создает новые вебинары, просматривает обратную связь и выгружает отчеты).

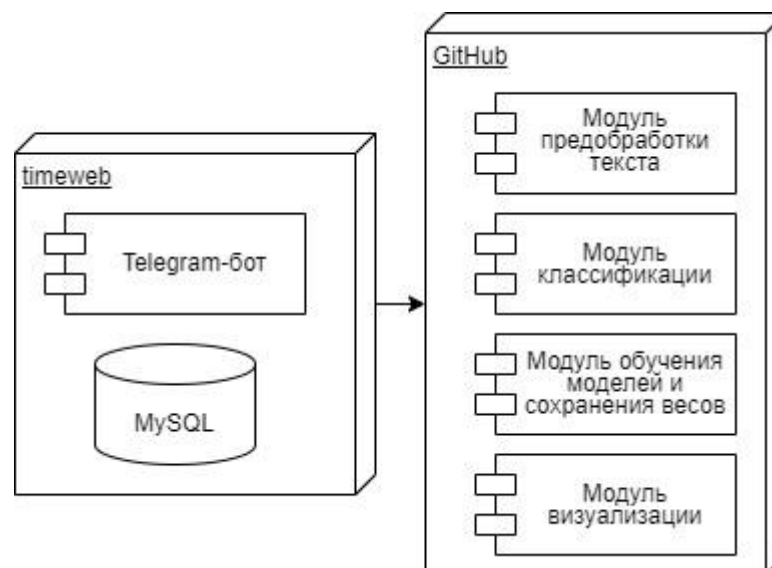
## Диаграмма активности



Возможно предусмотреть три варианты работы: работа Студента, работа Преподавателя, работа ML-инженера. В первом случае Студент только авторизуется и предоставляет обратную связь. Во втором случае Преподаватель может добавлять, как новые вебинары, так и получать агрегированную обратную связь от студентов, а также скачивать визуализации в Excel-формате. В третьем случае через какое-то время накопится новый пул данных, на основании которых ML-инженеру потребуется дообучить модель, чтобы метрики качества продолжали находиться на определенном уровне.

**Диаграмма компонентов**

Решение состоит из базы данных MySQL, ML-модуля (предобработка текста, обучение нейронных сетей решению задачи классификации текстов, визуализация), а также непосредственно модуль Telegram-бота.

**Диаграмма развертывания**

Для демонстрации результатов будет арендован сервер, где будет запущен Telegram-от. В рамках MVP не предусмотрено автоматическое дообучение моделей, однако в рамках масштабирования проекта возможна настройка CI/CD процесса дообучения моделей на GitHub Actions при достижении определенного количества новых опросов в базе данных.