

Интенсив №4

Команда:

Филиппов Дмитрий

Блинов Даниил

Цакаев Дени



самолет

Краткое содержание интенсива:



Содержание:

1. Постановка и описание задачи
2. Разделение работы
3. Работа с данными
4. Модели
5. Приложение
6. Самоанализ

Постановка и описание задачи

Название задания: Классификация позитивных комментариев по работе управляющей компании.

Метод выполнения:

Подготовка исходных данных.
Разработка модели классификации комментариев по работе управляющей компании на категории: Нравится скорость отработки заявок, Нравится качество выполнения заявки, Нравится качество работы сотрудников, Понравилось выполнение заявки, Вопрос решён, Другое.



Разделение работы

Внутри команды работа была распределена в соответствии с навыками и умениями каждого участника команды.

Обработка данных: Блинов Даниил.

Модель-трансформер: Блинов Даниил.

Модели и доп.обработчик: Филиппов Дмитрий.

Приложение-бот: Цакаев Дени.

Web-приложение: Филиппов Дмитрий.



Работа с данными

В ходе работы над данными была проведена работа по очистке данных и их анализу с выводением графиков для дальнейшей обработки данных для использования их в моделях.

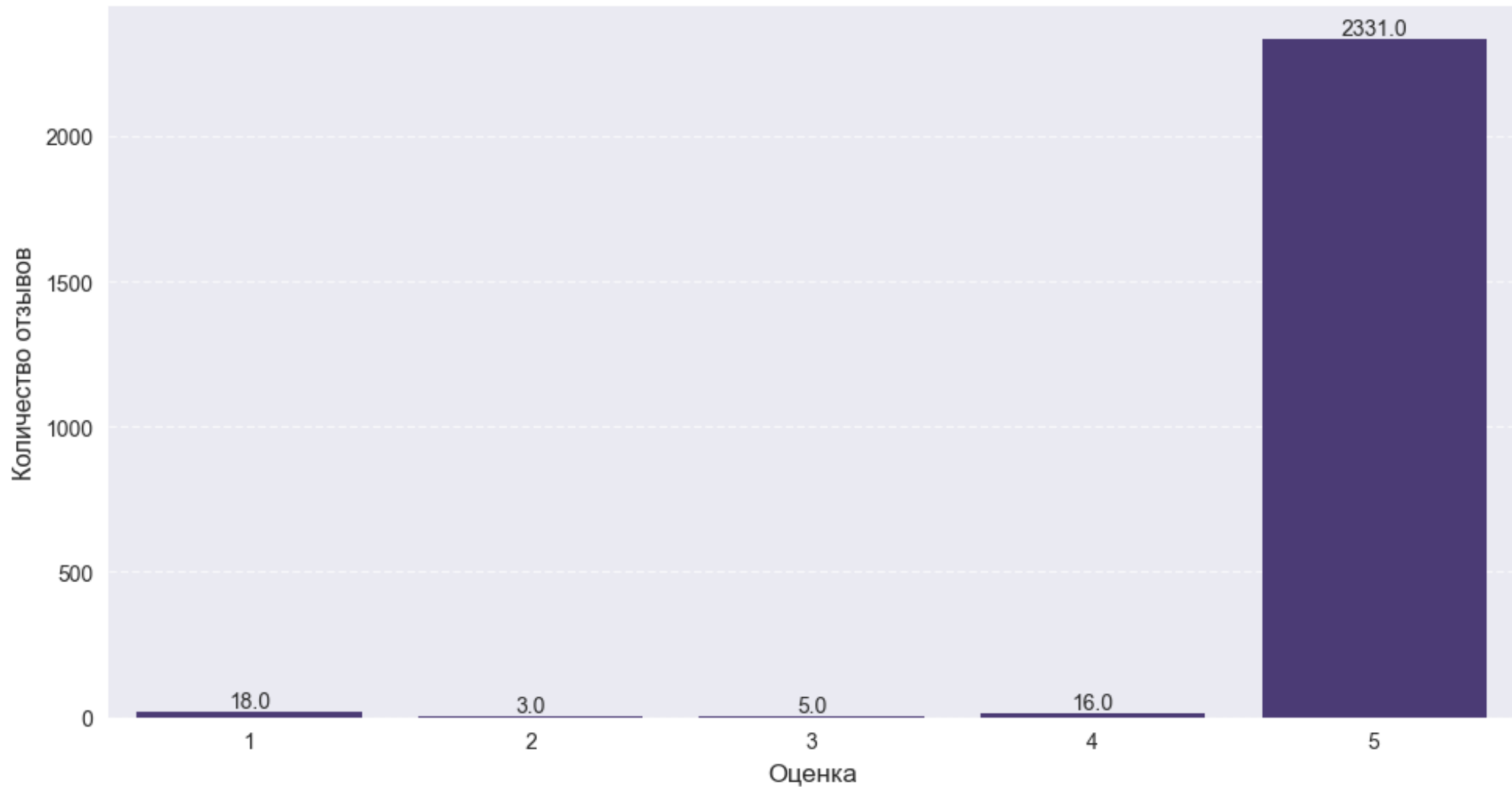
Результаты представлены ниже:

Также было разработано web-приложение и telegram-бот для классификации комментариев.



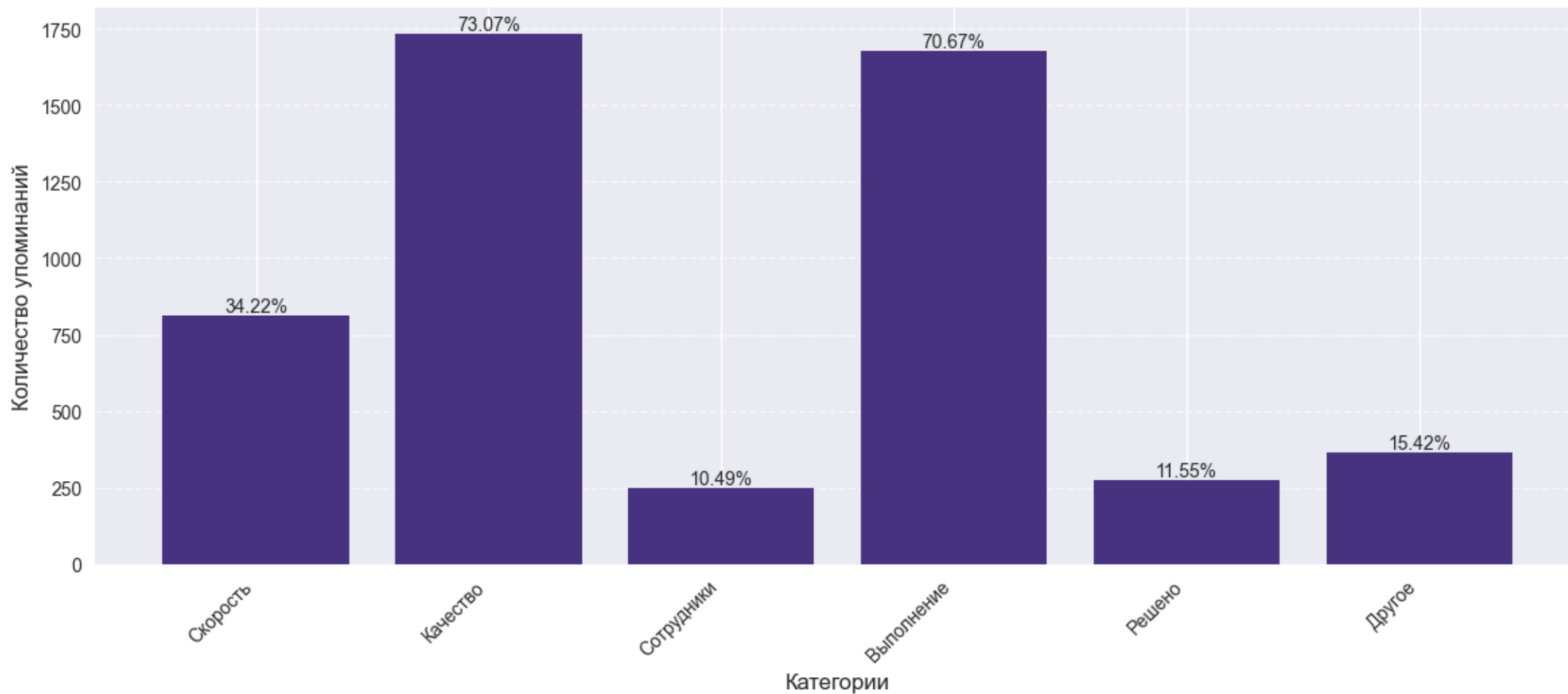
Работа с данными

Распределение оценок в отзывах

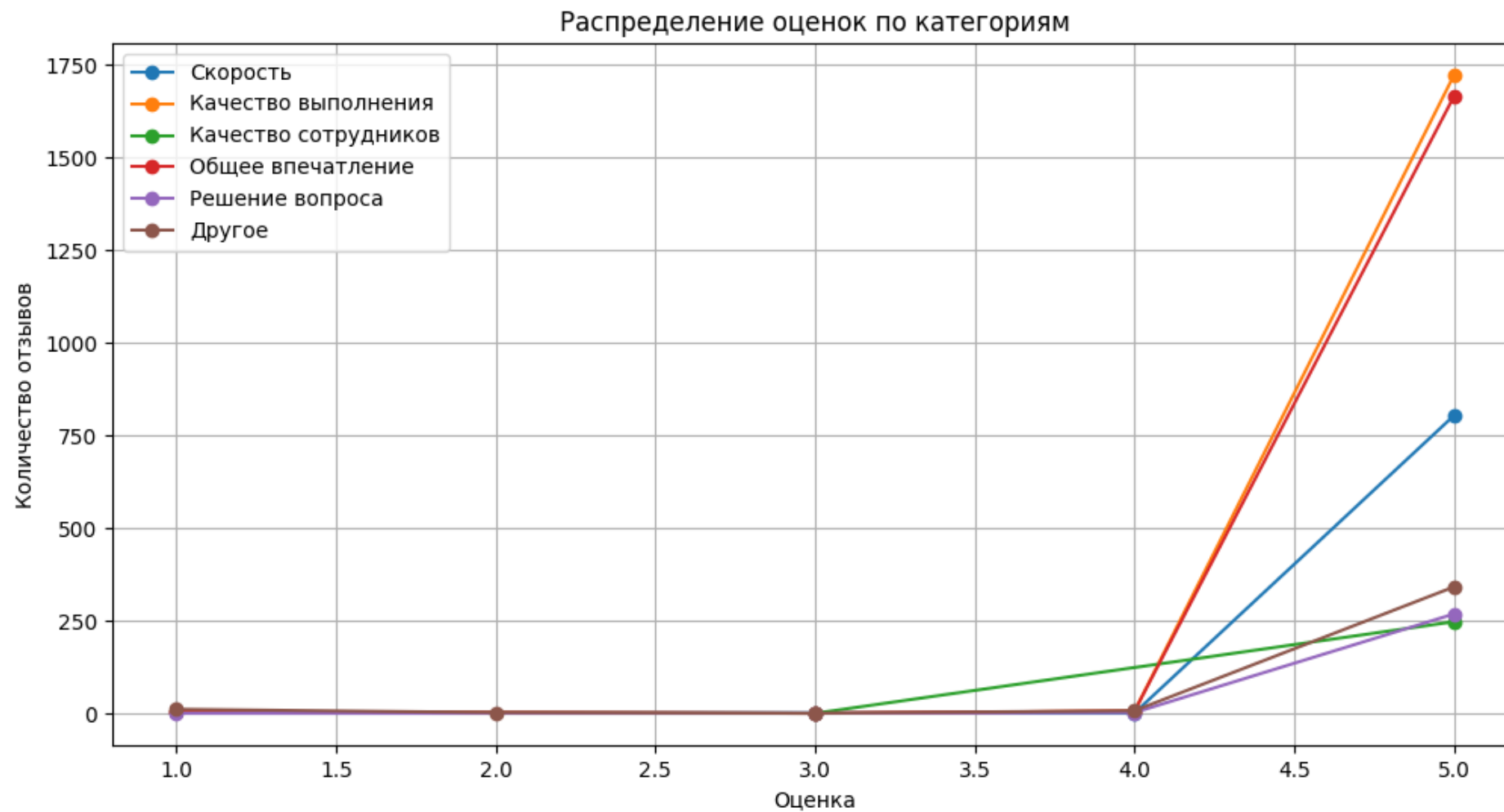


Работа с данными

Распределение отзывов по категориям

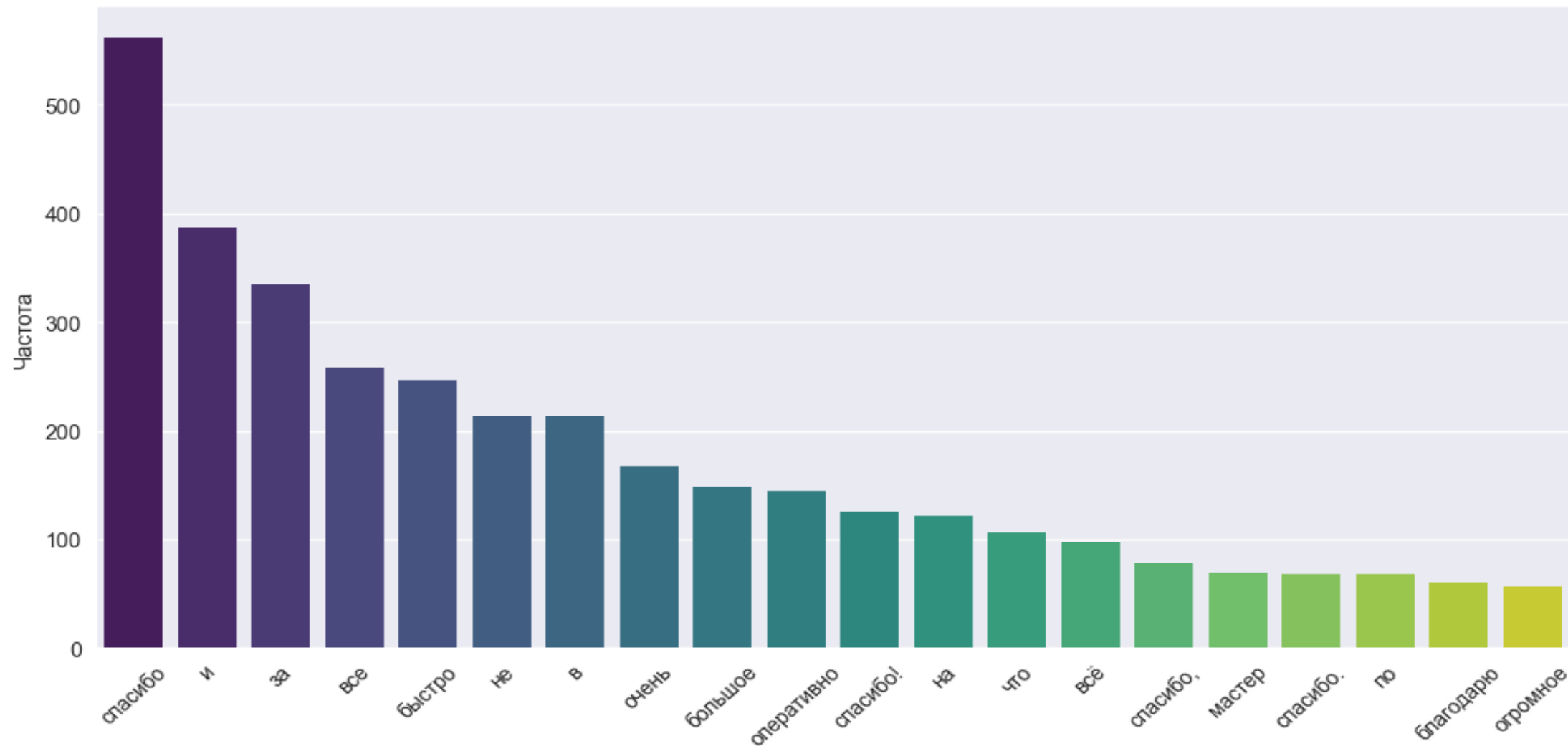


Работа с данными

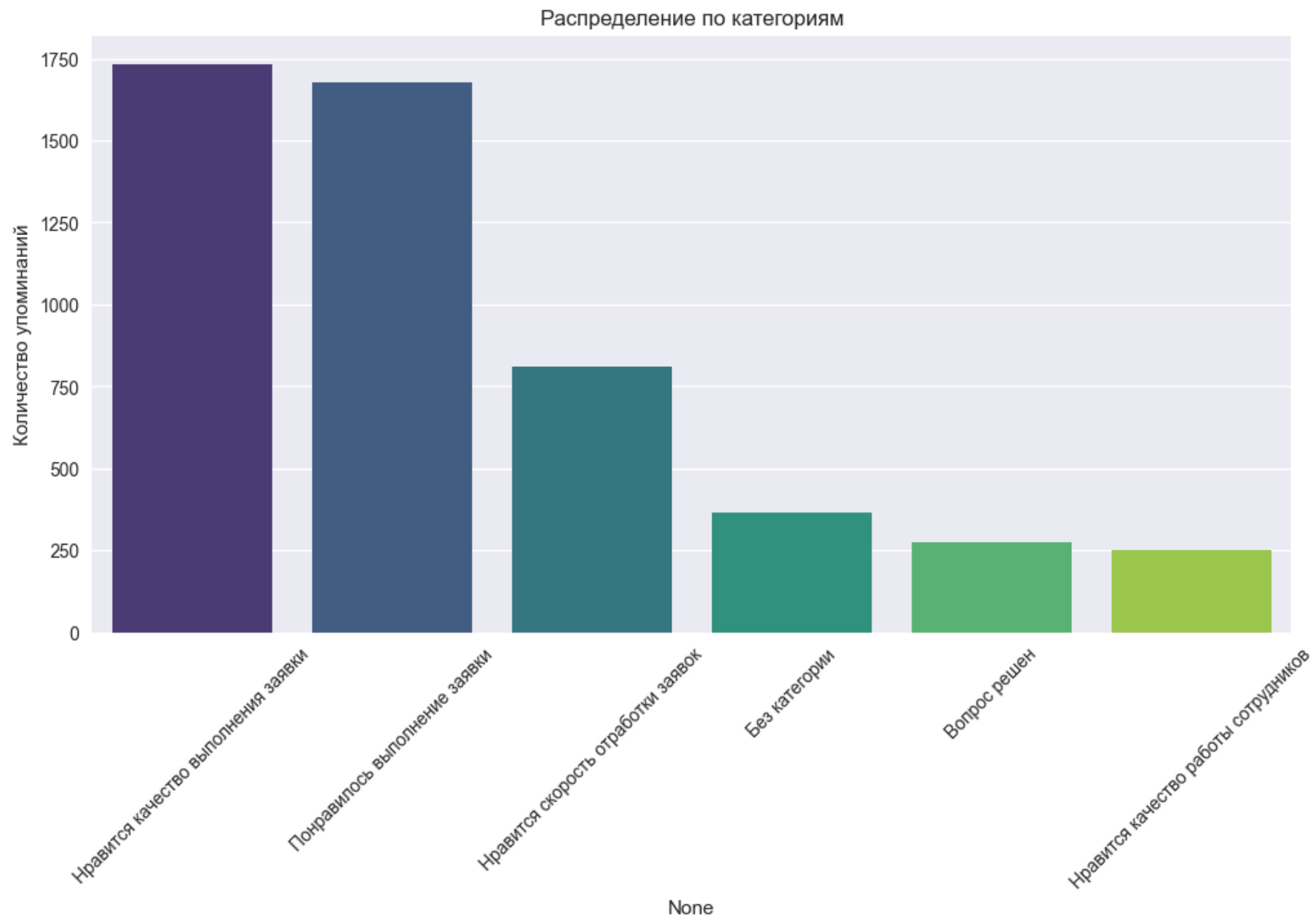


Работа с данными

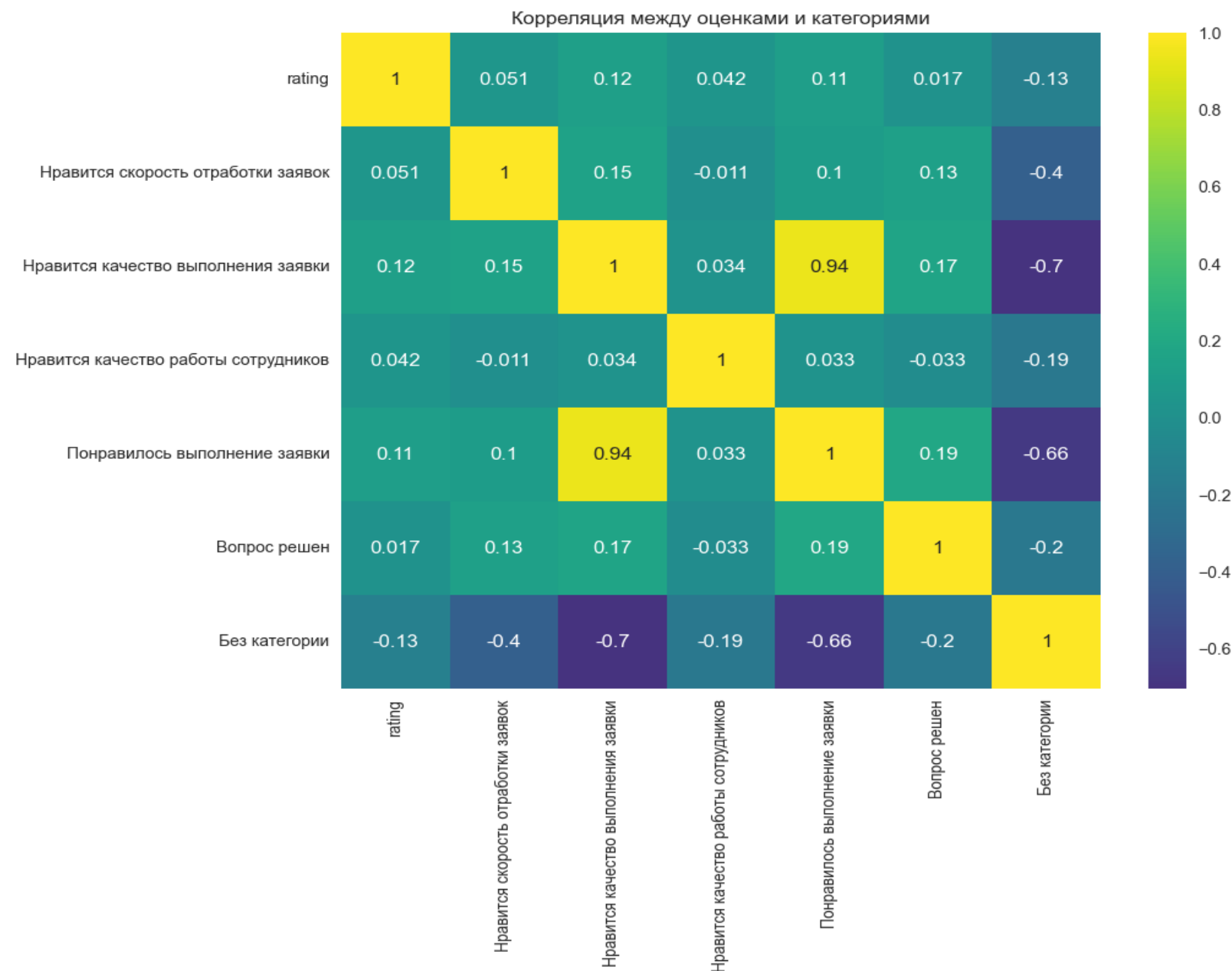
Топ-20 слов в положительных отзывах



Работа с данными



Работа с данными



Модели

Для выполнения поставленной задачи участниками команды использовались модели: XGBOOST, KNOWLEDGE GRAPH, LOGISTIC REGRESSION, RANDOM FOREST, BERT.

Ниже будут представлены результаты:



Модели

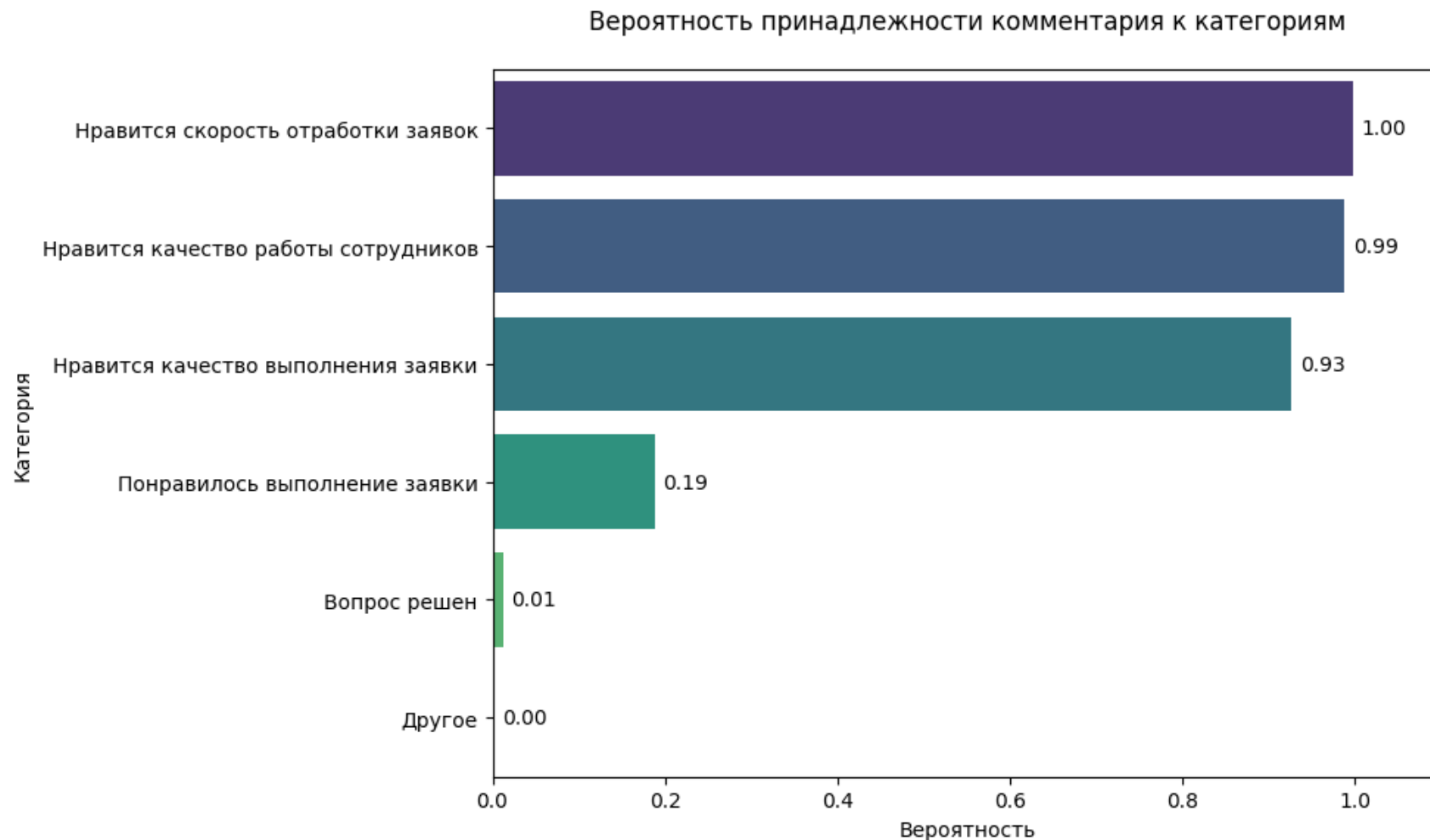
XGboost:

Пример: "Мастер

быстро пришёл

и качественно выполнил

работу".

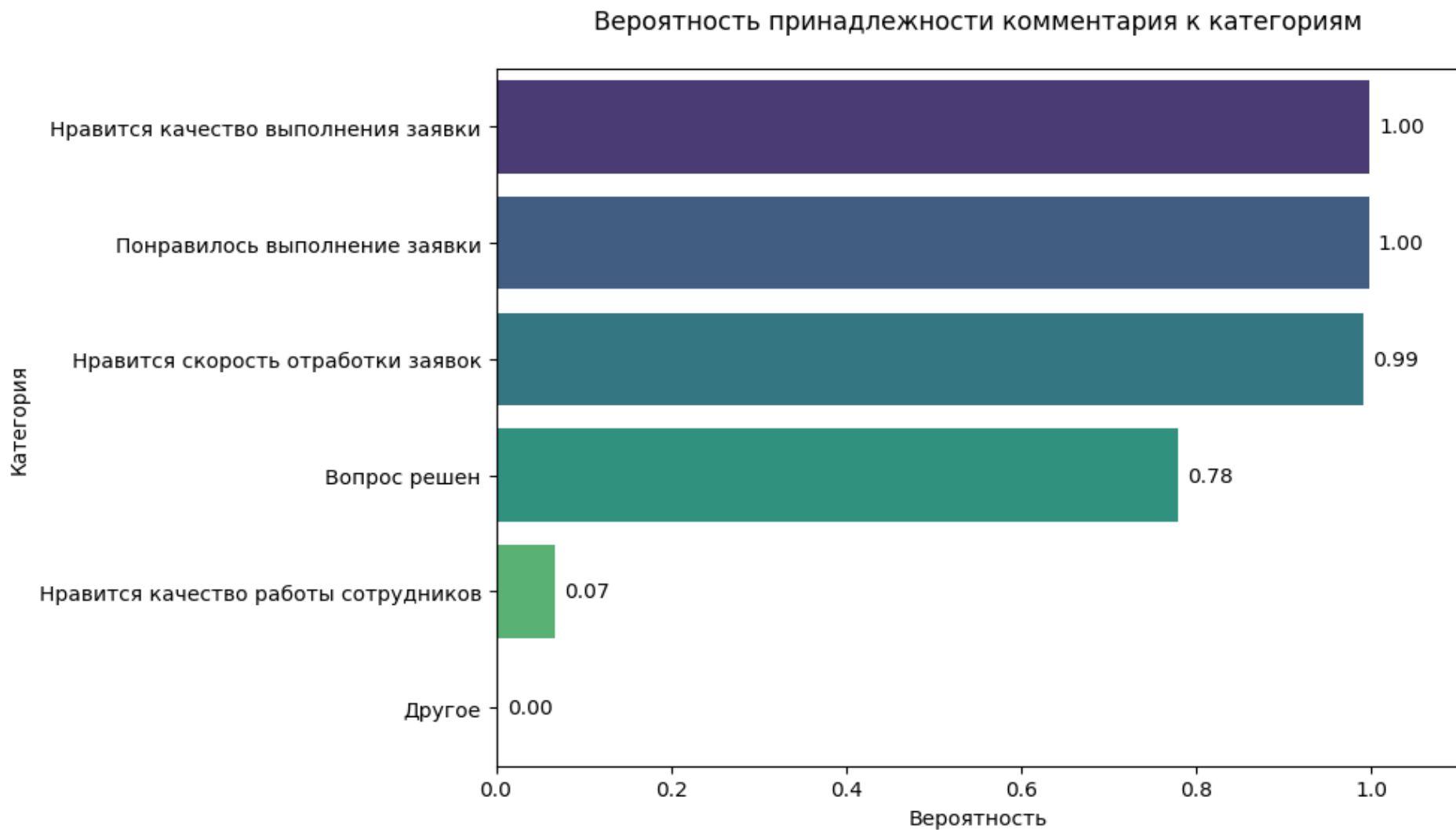


самолет

Модели

XGboost:

Пример: "Спасибо
за оперативность,
вопрос решен".

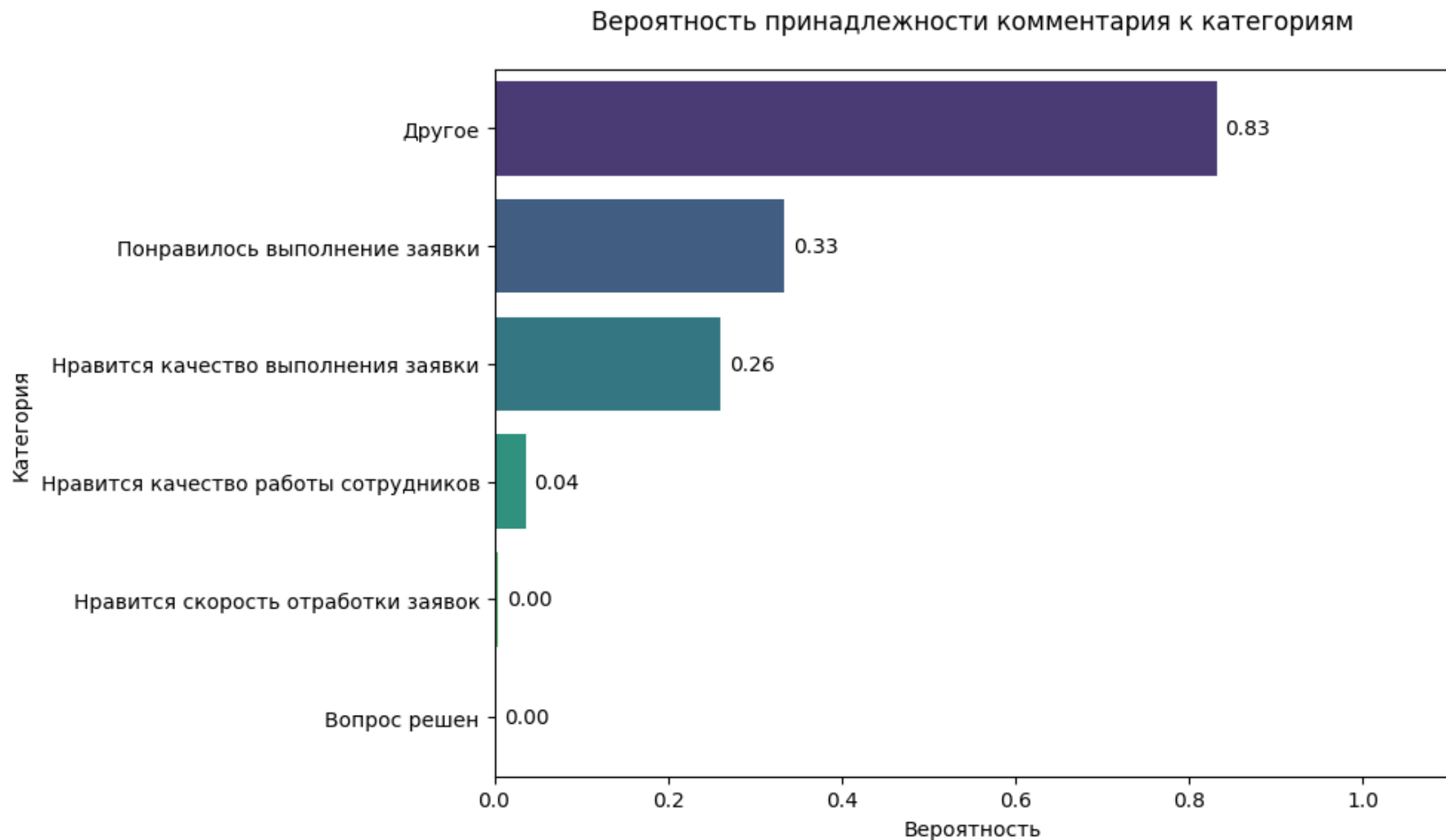


самолет

Модели

XGboost:

Пример: "Непонятно
что происходит,
никто не отвечает".



самолет

Модели

KNOWLEDGE GRAPH:

Это структурированное представление знаний в виде графа, где:

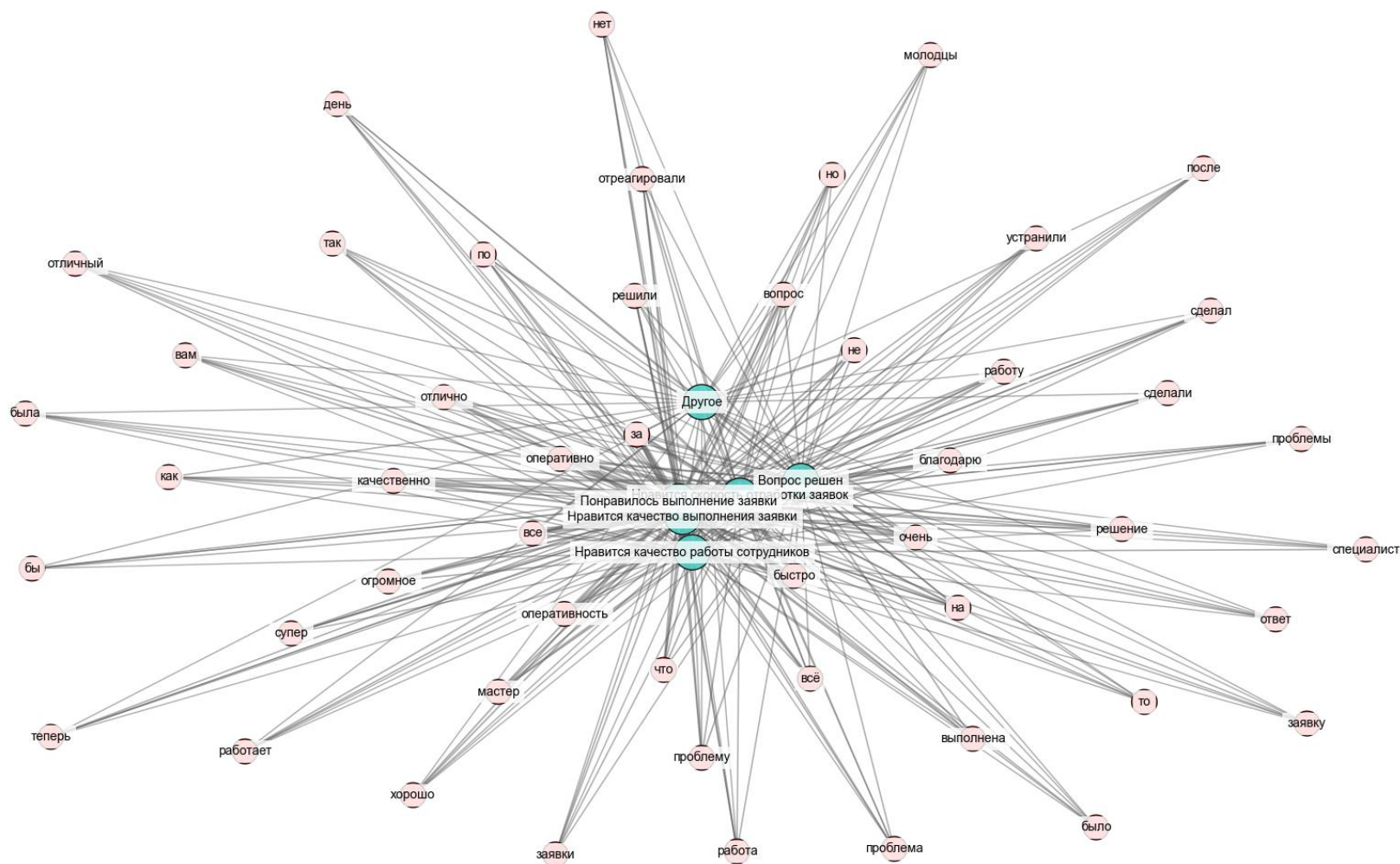
Узлы — это сущности (объекты, понятия, события).

Рёбра — это связи между ними (отношения).



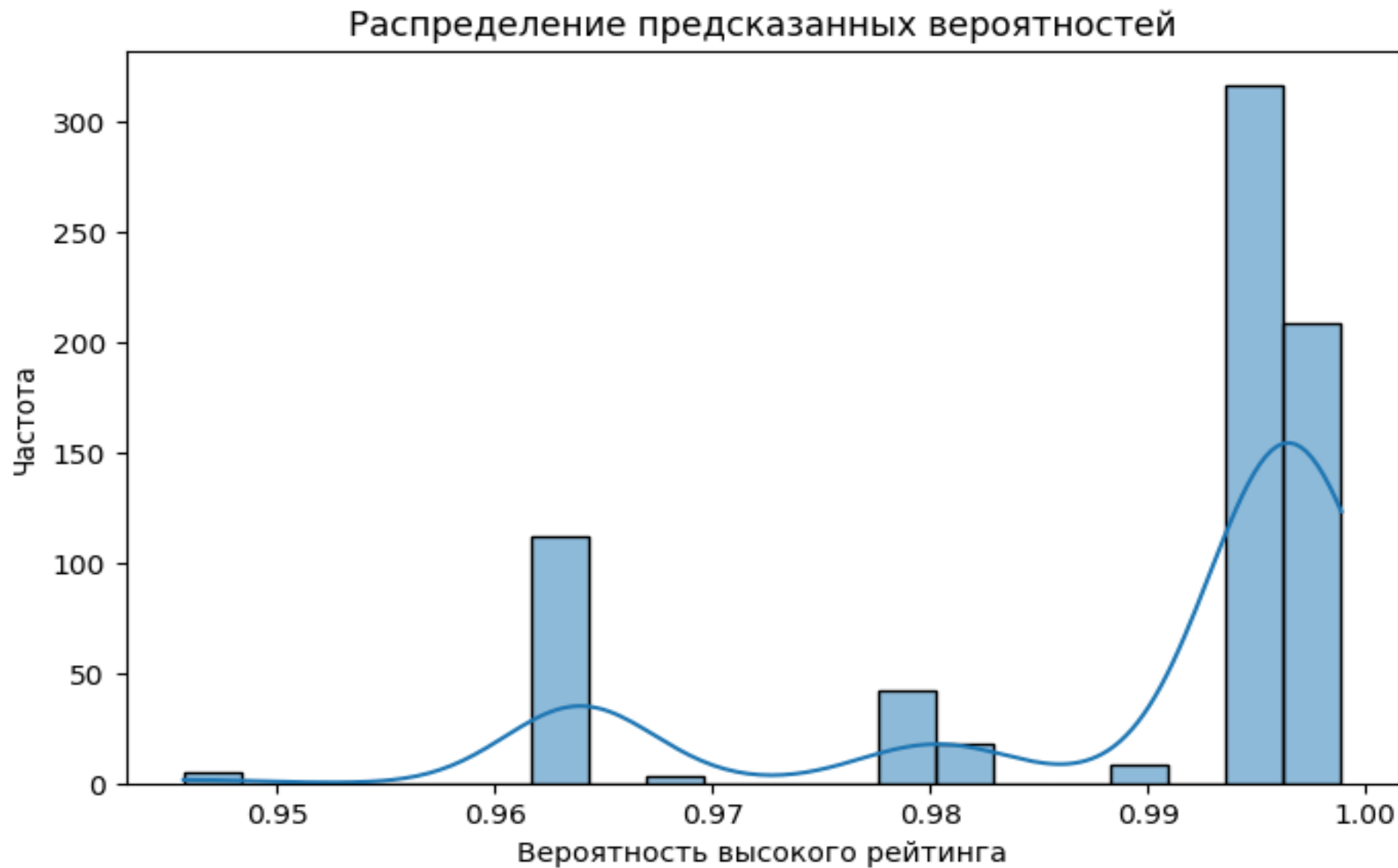
Модели

Граф знаний: Категории и ключевые слова



Модели

LOGISTIC REGRESSION:



самолет

Модели

В качестве модели-трансформера было решено использовать модель BERT из-за понимания контекста слов и возможности предобучения на двух задачах.

Результат представлен ниже:

Masked Language Model (MLM):

15% слов в предложении случайно маскируются, и модель учится их предсказывать.

Позволяет понимать глубокие связи между словами.

Next Sentence Prediction (NSP):

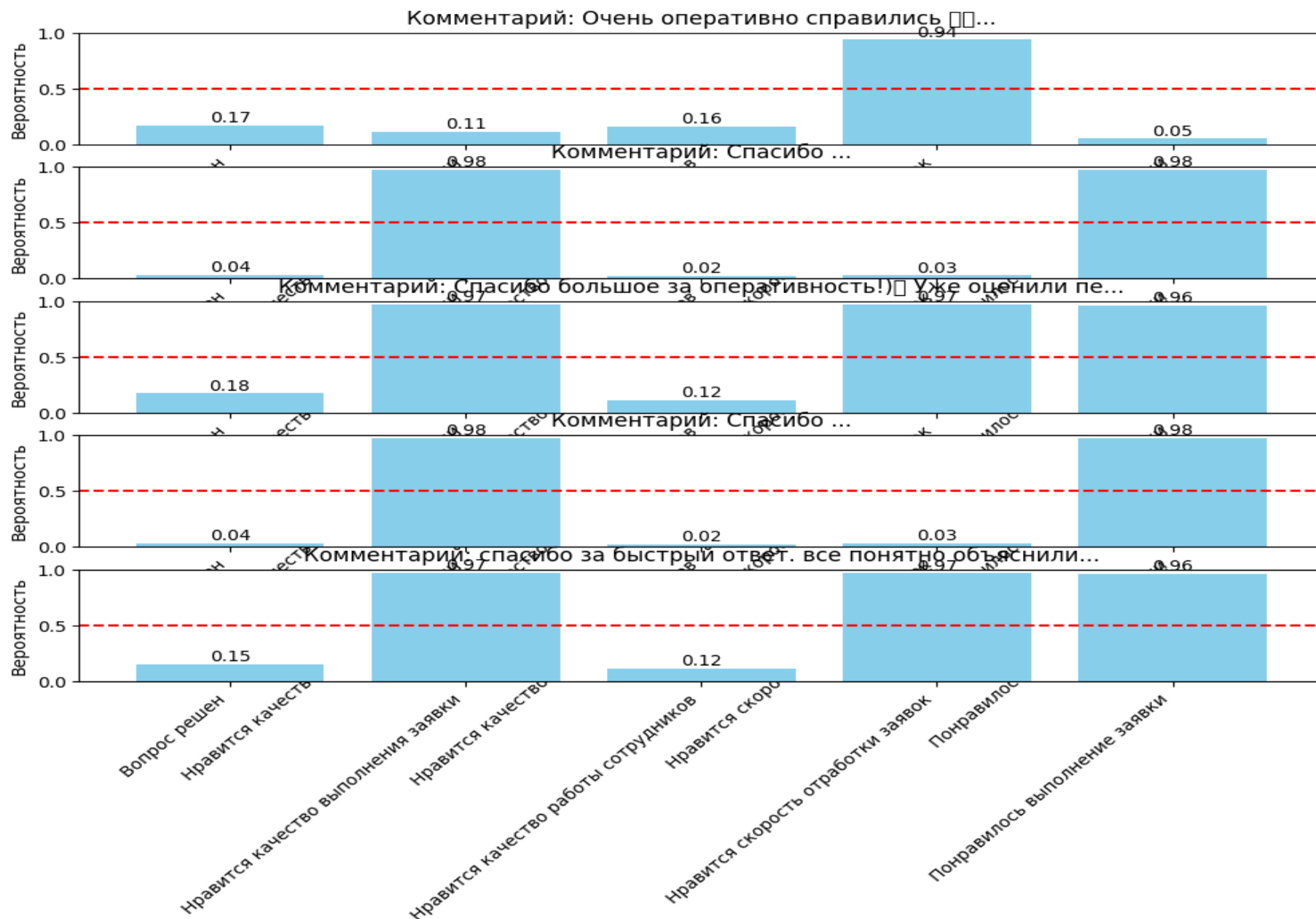
Модель определяет, следует ли одно предложение за другим в оригинальном тексте.

Помогает понимать отношения между предложениями.



Модели

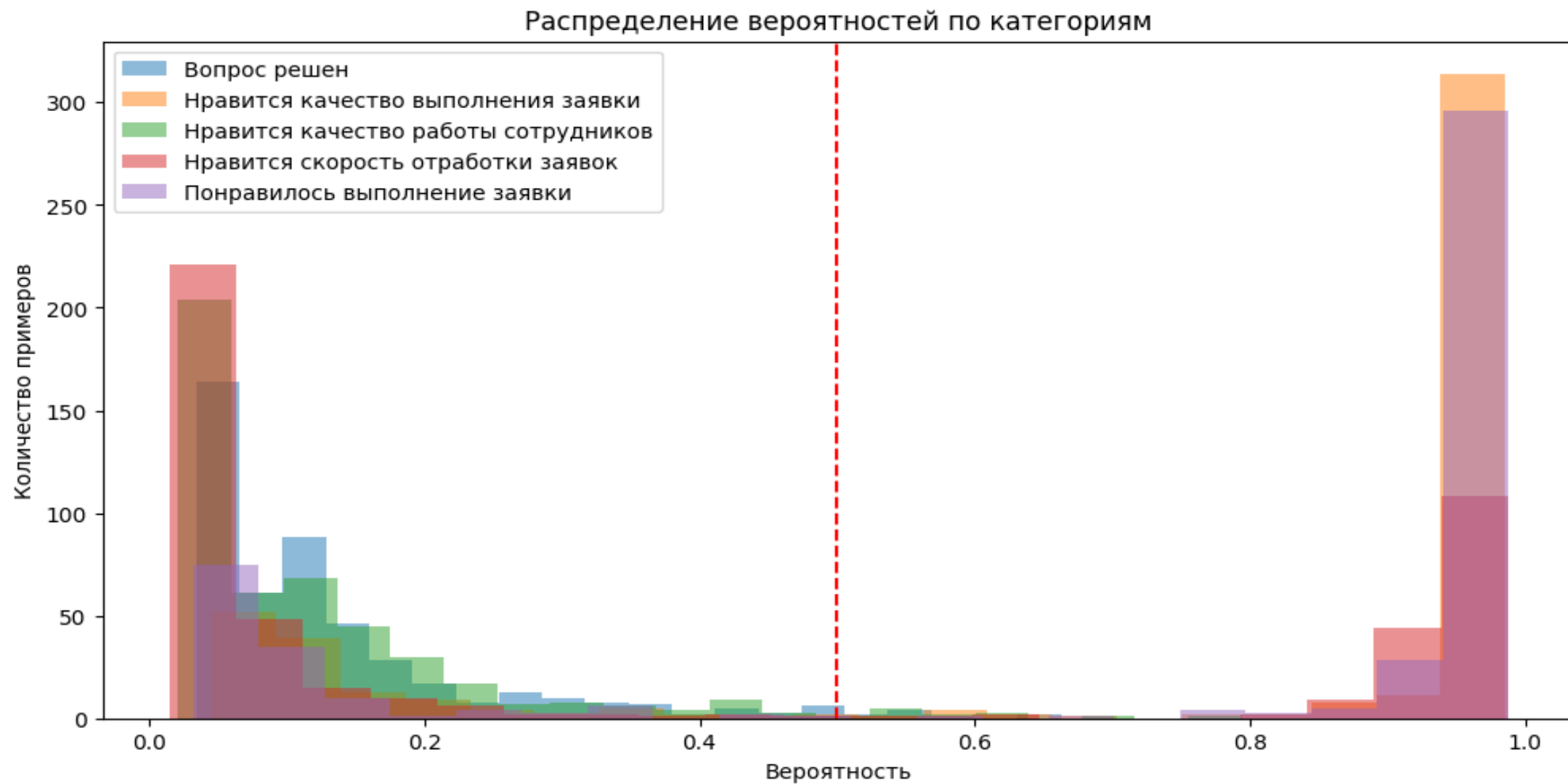
BERT:



самолет

Модели

BERT:



самолет

Модели

Модель Random Forest использовалась для реализации web-приложения по классификации комментариев.

Это ансамблевый алгоритм машинного обучения, основанный на **деревьях решений** (Decision Trees). Он используется для задач **классификации** и **регрессии**, отличается высокой точностью, устойчивостью к переобучению и способностью работать с большими объемами данных.



Приложение

Разработка web-приложения заняла значительное время и потребовала изучения теоретических материалов и применения их на практике.

Пользователь вводит комментарий, после чего нажимает на кнопку анализа или очистки (если необходимо убрать предыдущий результат)

В приложение загружается database, после запуска встроенная модель Random Forest обучается на загруженных данных (при повторном запуске она развёртывается сразу) и выводит процентную статистику распределения комментариев по категориям.

