

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Московский институт электроники и математики  
им. А.Н. Тихонова

Студенты учебной группы МСКМ-181:  
Горчавкина Анастасия Александровна  
Пехтерев Денис Олегович  
Чертенков Владислав Игоревич

Автореферат по проектной работе  
**Анализ отзывов о лекарственных препаратах в социальных  
медиа**

Научный руководитель к.ф.-м.н., доцент  
Артамонов С.Ю.

Москва 2019

## **1. Решаемая проектом проблема и ее значимость**

Проект относится к проблеме анализа информации, оставленной пользователями в Интернете о лекарственных препаратах и медицинской сфере в целом.

## **2. Поставленные в проекте цели и задачи**

Цель проекта состоит в применении современных средств прикладной математики для анализа данных.

Основные задачи проекта:

- Определение тональности отзывов методами машинного обучения
- Многоклассовая классификация отзывов
- Подбор гиперпараметров для эффективной реализации методов
- Визуализация данных на платформе Tableau Public

## **3. Насколько новы и работоспособны представленные методики и решения**

Для выполнения задач использованы следующие методы:

- Для определения тональности:
  - Метод логистической регрессии
- Для многоклассовой классификации:
  - Random Forest
  - Метод опорных векторов (SVM)
  - Градиентный бустинг над решающими деревьями

## **4. Персональный вклад каждого члена проектной команды**

Получены следующие результаты:

- Визуализация данных на платформе Tableau Public (Горчавкина А.А.)
- Разработана модель машинного обучения по определению тональности отзывов (Пехтерев Д.О.)
- Реализованы три метода многоклассовой классификации в среде Python:
  - Random Forest
  - Метод опорных векторов (SVM)
  - Градиентный бустинг над решающими деревьями (Чертенков В.И.)

## **5. Демонстрируемый результат проекта**

- Программный алгоритм, определяющий тональность текста, в основе которого лежит логистическая регрессия

- Программный алгоритм, классифицирующий тексты по соответствующим группам отзывов
- Визуализация в виде дашборда на платформе Tableau Public метрик качества тренировки модели машинного обучения

## **6. Апробация результатов**

Использованные методы справились с задачей с достаточной точностью.

## **7. Потенциал применения и внедрения**

Данный учебный проект имеет перспективы коммерческого применения. Итоговой продукт может заинтересовать как производителей лекарственных средств, так и площадок, предоставляющих аналитику для пользовательского потребления.