

## РАЗРАБОТКА МЕТОДА ОБНАРУЖЕНИЯ ДЕСТРУКТИВНОГО КОНТЕНТА В СОЦИАЛЬНЫХ СЕТЯХ С ПРИМЕНЕНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

**Шуваев Д.О.** (федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)  
**Научный руководитель – к.т.н., доцент (квалификационная категория «ординарный доцент») Менщиков А.А.** (федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

В данной работе рассмотрены методы машинного обучения, используемые для задач классификации информации. Исследована их эффективность в задачах определения противозаконных публикаций в социальных сетях. Предложена модель применения исследованных методов в совокупности для выявления деструктивного контента, представленного в текстовом и графическом виде.

**Введение.** За прошедшие несколько лет популярность социальных ресурсов существенно выросла благодаря тому, что все больше и больше пользователей стало обмениваться разнообразными видами информации их посредством. По статистическим данным в 2017 году в мире насчитывалось 2,8 млрд пользователей социальных сетей, а в январе 2022 года это значение увеличилось до 4,62 млрд. Таким образом, аудитория социальных медиа выросла на 65% (1,82 млрд уникальных пользователей) за 5 лет.

В связи с ростом общего количества пользователей различных социальных ресурсов растет и аудитория, которая подвергается воздействию деструктивного контента. Ввиду этого приобретает особую важность изучение публикуемого и, следовательно, распространяемого в социальных медиа интернет-контента.

**Основная часть.** Для решения поставленной задачи был проведен анализ эффективности применения популярных методов машинного обучения в задаче классификации неправомерного контента. Были исследованы такие методы, как Naïve Bayes, k Nearest Neighbors, Support Vector Machine, деревья решений и нейронные сети. На основе результатов работы данных методов были выделены те, которые успешнее справляются с классификацией текстовых материалов и те, которые лучше работают с графическими изображениями. Эффективность методов была проверена на метриках AUC/ROC-curve, F-score и Logistic loss.

**Выводы.** Таким образом, было произведено сравнение наиболее популярных методов машинного обучения в задаче классификации контента социальных сетей. Итогом работы является программное обеспечение, использующее сочетание наиболее эффективных методов машинного обучения. Разработанная программа позволяет анализировать публикации социальных сетей и давать положительную, нейтральную или отрицательную оценку касательно их содержания.

Шуваев Д.О. (автор)

Подпись

Менщиков А.А. (научный руководитель)

Подпись