АННОТАЦИЯ

диссертационной работы Сұлтан Данияра Рахманқұлұлы на тему: «Обнаружение и предотвращение кибербуллинга в онлайн-пользовательском контенте», представленной на соискание степени доктора философии (PhD) по образовательной программе «8D06301 – Системы информационной безопасности»

Актуальность темы исследования. Кибербуллинг - это растущая проблема, которая может иметь серьезные потенциальные проблемы как для жертв, так и для преступников. Это относится к использованию цифровых инструментов, таких как интернет-платформы, социальные сети и мобильные телефоны, для преследования, запугивания или другими способами причинения вреда людям. Кибербуллинг может принимать различные формы, включая публикацию злых или угрожающих сообщений, распространение слухов, обмен постыдным фото-, видео-контентом или исключение кого-либо из социальных групп онлайн.

Последствия кибербуллинга могут быть значительно пагубными. У пострадавших людей могут проявляться симптомы тревоги, депрессии, снижения самооценки и суицидальных мыслей. Кроме того, они могут столкнуться с проблемами со сном, аппетитом, концентрацией внимания и академическим или межличностным функционированием. В некоторых случаях кибербуллинг может привести к физическим последствиям, поскольку объект может испытывать изоляцию или воспринимать себя в опасности.

Люди, которые совершают кибербуллинг, также могут пострадать от негативных последствий. Они могут столкнуться с юридическими последствиями, если их действия будут признаны преступными, а также с социальными и профессиональными последствиями, такими как ущерб их репутации, трудности с приобретением друзей или поиском работы.

Важно уменьшить кибербуллинг для благополучия как жертв, так и комитетов. Есть несколько шагов, которые могут предпринять отдельные лица, школы и сообщества, чтобы избежать этой проблемы.

Одним из способов уменьшить кибербуллинг является информирование людей о последствиях их действий. Это можно сделать с помощью кампаний по информированию общественности, образовательных программ и ресурсов для родителей и учителей. Повышая осведомленность о последствиях кибербуллинга, люди, возможно, с большей вероятностью подумают, прежде чем прибегать к подобному поведению.

Еще одним способом уменьшить кибербуллинг является предоставление поддержки и ресурсов жертвам. Предоставляемая поддержка и ресурсы могут включать консультирование, терапию, а также механизмы для сообщения о случаях кибербуллинга и устранения их последствий. Благодаря этим инструментам жертвы могут чувствовать себя более уверенно и иметь возможность действовать и обращаться за помощью.

Также важно привлекать комитеты к ответственности за свои действия. Это может включать в себя дисциплинарное взыскание с учащихся, которые занимаются кибербуллингом, или возбуждение судебного иска в случаях, когда такое поведение считается преступным. Привлекая виновных к ответственности, это дает понять, что кибербуллинг недопустимо и может иметь серьезные последствия.

Цель диссертационной работы. Построение модели глубокой нейронной сети для автоматического обнаружения кибербуллинга в текстовых данных. Создание модели глубокого обучения для задачи бинарной классификации.

Задачи исследования.

1. Analysis of machine learning algorithms for binary and multiclass classification problems for cyberbullying detection.
2. Сбор данных и предварительная обработка данных на казахском языке для обучения алгоритмами машинного и глубокого обучения.
3. Анализ архитектур глубокого обучения. Обучающая реализация различных типов алгоритмов DL, таких как:
   a) Сверточные нейронные сети
   b) Глубокие нейронные сети
   c) Рекуррентные нейронные сети
   d) Сети с кратковременной долговременной памятью
   e) Многослойные персептроны
   f) Глубокие нейронные сети с использованием механизма внимания
   g) Проведение экспериментальных исследований, сравнение, выбор модели, настройка гиперпараметров для улучшения результатов модели.

Объект исследования: социальные сети (Vkontake, Instagram, Youtube, Twitter), новостные проталы(nur.kz, tengri news).

Предмет исследования: Алгоритмы машинного обучения и глубокого обучения для обнаружения кибербуллинга в текстовых данных.

Методы исследования: машинное обучение, глубокое обучение, теория нейронных сетей, data mining.

Научная новизна исследования:

– Разработана и обучена глубокая нейронная сеть с механизмом внимания для задачи двоичной и трех-классовой классификации в задаче выявления кибербуллинга.

– Создан датасет казахского языка, предварительно обработанный и помеченный вручную для задач машинного и глубокого обучения.

– Предложена новая схема нейронной сети, использующая механизм внимания в задаче классификации.

Теоретическая и практическая значимость работы. Теоретическая значимость работы заключается в исследовании существующих работ по выявлению кибербуллинга в текстовых данных, анализ инструментов обработки естественного языка. Практическая значимость исследовательской работы повышает точность алгоритмов глубокого обучения в задаче выявления кибербуллинга в онлайн медиа-пространстве. Результаты исследовательской работы опубликованы в международных научных журналах, индексируемых в базе данных SCOPUS и Web of Science, а также в изданиях, рекомендованных Комитетом по контролю в сфере образования и науки МНВО РК.

Главный вывод защиты. На основе механизма внимания, разработана новая глубокая нейронная сеть для выявления шаблоны кибербуллинга в текстовых данных. Эффективность предложенной модели была доказана экспериментально, приведя сравнительный анализ результатов предложенной модели с другими алгоритмами глубокого и машинного обучения.

Публикация результатов. В ходе проведения научного исследования по теме диссертации опубликовано 7 научных работ. Из них 4 статьи опубликованы в журналах, индексированных в базах Scopus и Web of Science, 1 статья – в изданиях, рекомендованных Комитетом по контролю в сфере образования и науки МНВО РК, 2 статьи в сборниках международных научно-практических конференций.

Объем и структура работы. Диссертациялық жұмыс 82 беттен тұрады және 34 сурет пен 15 кестеден тұрады. Мазмұны 6 бөлімнен тұрады.

Введение. В Данном разделе дано описание актуальности, новизны и основной цели диссертационной работы. Был приведен список основных задач и объекта и предмета исследования, а также теоретической и практической значимости исследования.

Первый раздел дает определение кибербуллинга, обзор на похожие проделанные работы в области выявления кибербуллинга. Представлен обзор результатов других авторов и инструментов обработки естественного языка.

Во втором заделе дается полное описание алгоритмов машинного обучения, теоритическое обоснование осоновных алгоритмов, а также их математическое обоснование. В разделе также подробно описана исследовательская часть работы, которая затрагивает бинарную и многоклассовую классификацию в текстовых данных для выявления кибербуллинга. Сводная информация приведена в виде графиков, таблиц и выводов.

В третьем разделе дается полное описание алгоритмов глубокого обучения, с приведением примеров и их строений. Также в разделе математически описаны все возможные виды слоев при создании нейронных сетей; структура механизма внимания разобрана с точки зрения математических формул и его использованием для классификации текста.

В четвертом разделе описаны проведенные работы по теме исследовательской диссертации. В первую очередь показан и разобран процесс создания парсера для сбора данных в социальных сетях и онлайн-пользовательском контенте, ручной классификации данных, первичной обработки и инструментов начальной обработки сырых данных, такие стемминг, лемматизация, удаление стоп слов и др. Далее приведены примеры использования алгоритмов машинного обучения в обработанных данных и готовых данных на примере использования набора данных Twitter. Далее приведены примеры и коды создания нейронных сетей. В конце раздела приведена предлагаемая модель с полным описанием вышеуказанных процессов.

В пятом разделе представлены результаты имплементированных алгоритмов машинного обучения и глубокого обучения. Приведены сводные результаты всех экспериментов. Также приведена таблица и диаграммы, показывающие прирост алгоритма долговременной кратковременной памяти с использованием механизма внимания по отношению к остальным методам.

В заключении обобщены практические результаты данной диссертационной работы, приведены ее самые значимые достижения в выявлении кибербуллинга в текстовом контенте с использованием алгоритмов машинного и глубокого обучения.