Detection and prevention cyberbullying in online user content
by
Sultan Daniyar Rakhmankululy
Dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy (Ph.D.) in the doctoral program
"8D06301- Information Security Systems"

ABSTRACT

Relevance of the research topic. Cyberbullying is a growing problem that can have serious potential problems for both victims and criminals. This refers to the use of digital tools, such as Internet platforms, social networks, and mobile phones, to harass, intimidate, or otherwise harm people. Cyberbullying can take various forms, including posting malicious or threatening messages, spreading rumors, sharing shameful photo or video content, or excluding someone from online social groups.

The consequences of cyberbullying can be significantly detrimental. Affected people may show symptoms of anxiety, depression, low self-esteem and suicidal thoughts. In addition, they may face problems with sleep, appetite, concentration, and academic or interpersonal functioning. In some cases, cyberbullying can lead to physical consequences, as the object may experience isolation or perceive itself in danger.

People who commit cyberbullying may also suffer from negative consequences. They may face legal consequences if their actions are found to be criminal, as well as social and professional consequences, such as damage to their reputation, difficulties in making friends or finding a job.

It is important to reduce cyberbullying for the well-being of both victims and committees. There are several steps that individuals, schools, and communities can take to avoid this problem.

One way to reduce cyberbullying is to inform people about the consequences of their actions. This can be done through public awareness campaigns, educational programs, and resources for parents and teachers. By raising awareness of the consequences of cyberbullying, people may be more likely to think before resorting to such behavior.

Another way to reduce cyberbullying is to provide support and resources to victims. The support and resources provided may include counseling, therapy, as well as mechanisms for reporting cases of cyberbullying and addressing their consequences. Thanks to these tools, victims can feel more confident and be able to act and seek help.

It is also important to hold committees accountable for their actions. This may include disciplinary action against students who engage in cyberbullying, or legal action in cases where such behavior is considered criminal. By bringing the perpetrators to justice, it makes it clear that cyberbullying is unacceptable and can have serious consequences.

The purpose of the dissertation work. Building a deep neural network model for automatic detection of cyberbullying in text data. Creating a deep learning model for a binary classification problem.

Research objectives:

- Analysis of machine learning algorithms for binary and multiclass classification problems for cyberbullying detection.
- Data collection and pre-processing of data in Kazakh for teaching ML and DL algorithms to complete a thesis.

Analysis of deep learning architectures. Training implementation of various types of DL algorithms, such as:

a) Convolutional neural networks
b) Deep neural networks
c) Recurrent neural networks
d) Networks with short-term long-term memory
e) Multilayer perceptrons
f) Deep neural networks using the attention mechanism
g) Conducting experimental studies, comparison, model selection, setting up hyperparameters to improve the results of the model.

The object of study. social networks (Vkontakte, Instagram, Youtube, Twitter), news channels (nur.kz , tengri news).

The subject of study. Machine learning and deep learning algorithms for detecting cyberbullying in text data.

Research methods: machine learning, deep learning, neural network theory, data mining.

The scientific novelty of the research:

– A deep neural network with an attention mechanism has been developed and trained for the task of binary and three-class classification in the task of detecting cyberbullying.

– A dataset of the Kazakh language has been created, pre-processed and marked manually for machine and deep learning tasks.

– A new neural network scheme using the attention mechanism in the classification problem is proposed.

The theoretical and practical significance of the work. The theoretical significance of the work lies in the study of existing works on the identification of cyberbullying in text data, analysis of natural language processing tools. The practical significance of the research work increases the accuracy of deep learning algorithms in the task of detecting cyberbullying in the online media space. The results of the research work have been published in international scientific journals indexed in the SCOPUS database and Web of Science, as well as in publications recommended by the Committee for Control in the Field of Education and Science of the Ministry of Education and Science of the Republic of Kazakhstan.

The main conclusion of the defense. Based on the attention mechanism, a new deep neural network has been developed to identify cyberbullying patterns in text data. The effectiveness of the proposed model was proved experimentally by providing a comparative analysis of the results of the proposed model with other deep and machine learning algorithms.

Publication of results. During the scientific research on the topic of the dissertation, 7 scientific papers were published. Of these, 4 articles have been published in journals indexed in Scopus and Web of Science databases, 1 article in publications recommended by the Committee for Control in the Field of Education and Science of the Ministry of Education and Science of the Republic of Kazakhstan, 2 articles in collections of international scientific and practical conferences.

The volume and structure of the work. The dissertation work consists of 82 pages and includes 34 figures and 15 tables. The content includes 6 sections.

The introduction section describes the relevance, novelty and main purpose of the dissertation work. A list of the main tasks and the object and subject of the study, as well as the theoretical and practical significance of the study was given.

The first section gives a definition of cyberbullying, an overview of similar work done in the field of cyberbullying detection. An overview of the results of other authors and natural language processing tools is presented.

The second section a full description of deep learning algorithms is given, with examples and their structures. Also, the section mathematically describes all kinds of layers when creating neural networks; the structure of the attention mechanism is analyzed in terms of mathematical formulas and its use for text classification.

The third section a full description of deep learning algorithms is given, with examples and their structures. Also, the section mathematically describes all kinds of layers when creating neural networks; the structure of the attention mechanism is analyzed in terms of mathematical formulas and its use for text classification.

In the fourth section the work carried out on the topic of the research dissertation is described. First of all, the process of creating a parser for collecting data in social networks and online user content, manual data classification, primary processing and tools for initial processing of raw data, such as stemming, lemmatization, removal of stop words, etc., is shown and analyzed. The following are examples of using machine learning algorithms in processed data and ready-made data using the Twitter dataset as an example. The following are examples and codes for creating neural networks. At the end of the section, the proposed model is given with a full description of the above processes.

The fifth section presents the results of implemented machine learning and deep learning algorithms. Summary results of all experiments are given. There is also a table and diagrams showing the growth of the algorithm of long-term short-term memory using the attention mechanism in relation to other methods.

In conclusion the practical results of this dissertation work are summarized, its most significant achievements in identifying cyberbullying in text content using machine and deep learning algorithms are presented.