Journal of Theoretical and Applied Information Technology
15th April 2022. Vol.100. No 7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AUTOMATED DETECTION OF DESTRUCTIVE CONTENTS ON THE INTERNET USING DATA MINING AND MACHINE LEARNING METHODS

[1]RYSKUL MEDETBEKOVA, [1]ZLIKHA MAKHANOVA, [2]BAYAN MYRZAKHMETOVA, [2,3]AIGERIM TOKTAROVA, [1]AIGUL IMANBAYEVA, [4]KADRZHAN SHIYAPOV, [5]GULNARA SEIDALIYEVA, [6]GAUKHAR SEIDALIYEVA, [7]NURBEK KONYRBAEV, [8]RUSTAM ABDRAKHMANOV

[1]M.Auezov South Kazakhstan University, Shymkent, Kazakhstan
[2]South Kazakhstan State Pedagogical University, Shymkent, Kazakhstan
[3]Khoja Akhmet Yassawi International Kazakh Turkish University, Turkistan, Kazakhstan
[4]Abai Kazakh National Pedagogical University, Almaty, Kazakhstan
[5]Kazakh National Agrarian Research University, Almaty, Kazakhstan
[6]University of International Business, Almaty, Kazakhstan
[7]Korkyt Ata Kyzylorda Ubiversity, Kyzylorda, Kazakhstan
[8]International University of Tourism and Hospitality, Turkistan, Kazakhstan
E-mail: aikerimtoktarova@gmail.com

ABSTRACT

In the context of ensuring the information and psychological security of society, the problem of uncontrolled growth of destructive content on the Internet is considered. Solving the problem of searching and identifying destructive information requires a minimum degree of subjectivity and maximum automation, since currently it is solved mainly by expert methods with the compilation of registers of prohibited sources. The existing search methods (expert processing, thematic search, intelligent data processing methods) used to find specific words in the text are considered, the advantages and disadvantages of these methods are noted. To solve this problem, methods of data mining and machine learning are proposed to search for destructive content using the example of profanity in textual information. The proposed approach is distinguished by the possibility of self-completion of the dictionary by a system based on the judgment of the identification of an unfamiliar word classified as non-normative. The correctness of the dictionary replenishment is measured using different metrics. A detailed description of this study is presented, starting from data collection to classification and detection of destructive content. An example of the algorithm and the initial results of its testing are given.

Keywords: *Destructive Data, Security, Internet, Machine Learning, Detection, Natural Language Processing.*

1. INTRODUCTION

The progress of science and technology is currently accompanied by the intensive introduction of new information technologies into many spheres of human activity. The development of the Internet leads to an uncontrolled exponential increase in the amount of various information, most of it presented in text form [1].

To ensure information security, it is of great importance to analyze content in telecommunications networks containing illegal information, including data related to terrorism, drug trafficking, preparation of protest movements or mass riots, containing offensive statements against state symbols, profanity, etc. [2-3]. In this regard, as well as in the context of the potential creation of a "closed" Internet space, solving the problem of searching and automatic identification of destructive information is a priority. At the same time, to date, legislative measures are insufficient to ensure the information and psychological security of society [4].

Thus, there is a need to develop specialized search systems and categorization of information resources. In addition, due to the ever-growing volume of information resources, automation of the process of identifying the nature of input text information in order to further block dangerous content will not only reduce labor costs, but also minimize subjectivity and the likelihood of errors due to the influence of the human factor [5].

We will pay special attention to relatively new external threats associated with the active development of social networks, social engineering tools, as well as the "criminalization" of the Internet.

There are the following ways to protect information [6-9]:

- threat prevention — preventive measures to ensure information security in the interests of anticipating the possibility of their occurrence;
- the identification of threats is expressed in the systematic analysis and control of the possibility of the appearance of real or potential threats and timely measures to prevent them;
- threat detection aims to identify real threats and specific criminal activities;
- localization of criminal actions and taking measures to eliminate the threat or specific criminal actions;
- elimination of the consequences of threats and criminal actions and restoration of the status quo.

Conditionally, external threats can be divided into "traditional" and "new". "Traditional" external threats include spam, phishing, computer viruses, Trojans and network attacks [10]. The next chapter will deal with "new" external threats.

2. "NEW" EXTERNAL THREATS

Indeed, the topic of Internet communication is relevant for our time. It is impossible to answer unequivocally what the development of Internet technologies has led to to a greater extent. An international network, usually abbreviated to the Internet, simply means a global network of millions of computer networks. In other words, it is a network of all computer networks around the world or just a network of networks. The Internet allows you to instantly exchange information anywhere in the world. Through e-mail, often referred to as email for short messages, instant messaging, or chats, the Internet allows people to communicate with other people on both an interpersonal and mass communication level. It also provides its users with access to the volumes of information available on the World Wide Web [11]. Data larger than any known encyclopedia can be transferred from one part of the world to the farthest part of the globe at the speed of light. In this way, users can upload or download information at an astounding speed. Figure I demonstrates the number of active users involved in social networks (in millions) [12].

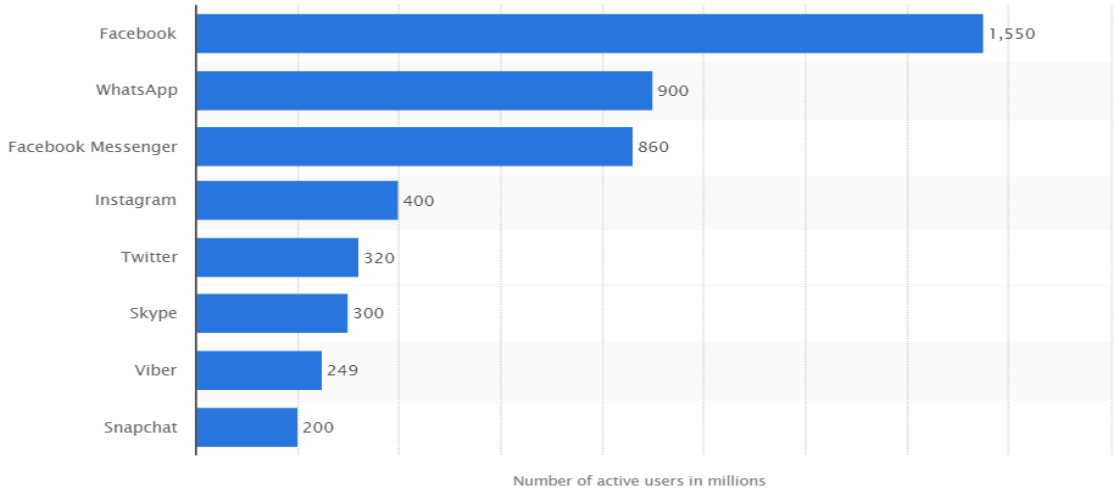

Figure 1: Number Of Active Users (In Millions) On Various Social Networking Websites

Technology-based and content-based cybercrime are the two types of cybercrime. The content-based crime is carried out by any terrorist organization linked to sexual harassment, fear, child pornography, national security, and so on. Hacking, spying, and malicious code injection are examples of technology-based criminality [13]. The taxonomy of cybercrime is shown in Figure II, along with several instances. People who fall into both groups should be aware of their surroundings. Cyber criminals tend to dwell in many parts of the world and enjoy being honored by a variety of countries.

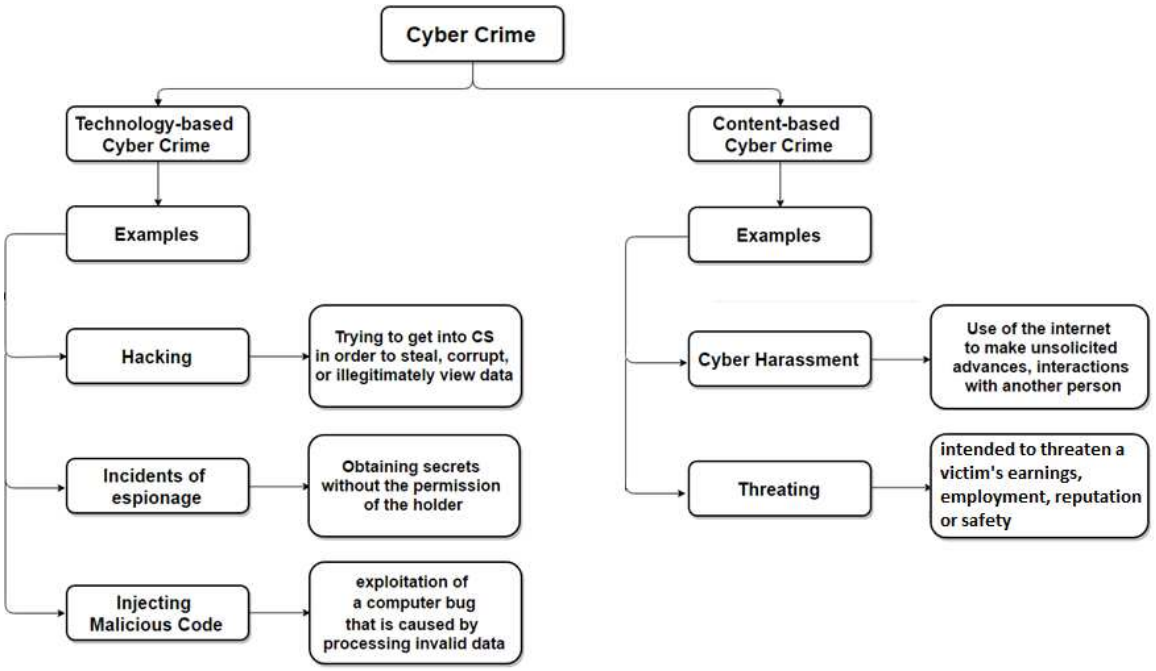

Figure 2: Taxonomy of cybercrime with examples [14]

2.1. Ransomware programs

The wave of distribution of such programs overwhelmed the domestic segment of the Internet just a few years ago. These programs block the operation of the operating system and/or encrypt all files on the hard disk [15]. After the blocking or encryption process is completed, the user is asked to transfer a certain amount (usually in the region of 10-20 thousand rubles) to receive an unlock / decryption code. As a confirmation of the ability to decrypt files, the attackers offer to send them 2-3 encrypted files, and then return them in their initial form. The threat is fraught with the fact that when the operating system is blocked, it can simply be reinstalled, but in the case of encrypting all files, this will not help.

During the outbreak of the ransomware epidemic (2012-2013), antivirus companies offered unblocking services. The work of the services was based on the vulnerabilities of the ransomware code and the possibility of selecting an unlock code based on information about the behavior of the program and the requests issued by it. However, since then, these malicious programs have been significantly improved and equipped with anti-code analysis tools, so the success of using antivirus services has significantly decreased (it only works when infected with "old" malicious software code) [19].

2.2. Unwanted content

Such content should include any information that is unpleasant for the user and does not correspond to his request, since many sites display the corresponding pop-up windows [20]. It can be intrusive advertising, pornography, an offer to play at an online casino or use some paid online service, etc. Also, unwanted content may contain sites found by a search engine query. For example, if a child enters the query "Little Red Riding Hood", then the first links issued by Yandex will be associated with a 2011 film belonging to the category "horror, fantasy, thriller".

Separate types of unwanted content can be considered flood, offtop, flame and holivar [21]. Flood - posting unnecessary, meaningless information on forums and in chat rooms. Offtop - posting information on forums and in chat rooms that is not related to the topic under discussion, such messages are called "offtopic" from the English. off — from, away and topic — topic, question, i.e. literally — off topic, question. Flame from the English flame — flame, fervor, passion - the exchange of messages in Internet forums and chat rooms, which is a war of words, often no longer relevant to the original cause of the dispute [21]. Flame is often an excuse or a means for trolling, as will be discussed below. Holivar from the English holy war is a holy war — an exchange of messages in Internet forums and chat rooms,

which is a meaningless discussion in which participants try to prove to each other the advantage of one of several similar alternatives (computer programs, technologies, actors, musical groups, etc.) [20].

2.3. Internet fraud

These threats include all types of traditional fraud that have successfully moved into the virtual space. As a rule, users are faced with the following:

1. "Letters of happiness" are messages by e-mail or on a social network containing information that the addressee has become the owner of a significant prize, or that a certain benefactor or a suddenly found distant relative wants to transfer a certain amount to him. To receive a "win" or "donation", it is proposed to transfer a certain amount to the appropriate account.
2. "Help the Child— - fundraising for a seriously ill child or other charitable purposes. Often, attackers use the data of a real person for whose needs a fundraiser is organized, but at the same time they substitute the details for transferring donations.
3. Payment for access to some, as a rule, illegal or semi-legal resource, for example, with pirated content. Different payment methods may be offered, but after it is carried out, no access is granted to the user, and in the case of payment by SMS, the user, without knowing it, may find himself subscribed to a service unknown to him, for which funds will be regularly debited from his mobile phone account.
4. Offer to participate in various projects that require payment of a fee to join.
5. The sale of goods with a prepayment condition, if in reality the goods do not arrive to the buyer (or a cheaper thing arrives).

2.4. Trolling and cyberbullying

Trolling on the Internet is any aggressive actions, such as direct insults to other users, publication of provocative or offensive messages, gross violation of the norms and rules of communication. Cyberbullying is an aggressive, intentional act committed by a group of persons or one person using electronic forms of contact, repeated repeatedly and prolonged over time against a victim who finds it difficult to protect himself [22].

Foreign psychologists have defined the situation of violence against a person in an educational environment - bullying (school bullying). The problem of bullying abroad was developed by such authors as S. M. Agogo, K. Lee, E. Roland, etc. This phenomenon has been studied not so long ago, but a list of basic measures to overcome bullying has already been developed, psychological mechanisms have been revealed and personal characteristics of bullying participants have been given (I. B. Achitaeva, I. S. Cohn, E. I. Feinstein, etc.).

With the development of information technology, significant changes have taken place in the life of a modern teenager: a virtual reality has appeared in which communication and interpersonal relationships are moving to a new, unfamiliar level for them. Bullying becomes more dangerous for an individual, as it can be carried out with the help of Internet technologies.

For the first time, the definition of "cyberbullying" was given by Bill Belsey. In his opinion, cyberbullying is the use of information and communication technologies, for example, e-mail, mobile phone, personal Internet sites, for intentional, repeated and hostile behavior of a person or group aimed at insulting other people [23].

Based on the definition of cyberbullying, its main difference from traditional bowling can be distinguished: all actions directed against a person take place in a virtual space. But this is not the only difference. In the virtual space, it becomes possible to create an alternative "self-image", as a result of which a deformation of the real "self-image" may occur; at the same time, the teenager is not responsible for his actions.

In practice, cyberbullying can be:

- systematically sending abusive messages to the victim's email address or posting them on the victim's social network page;
- creation of a "hate wall" page on the social network on which a group of cyberbullers posts offensive messages against the victim;
- posting personal photos, screenshots of correspondence on a social network or on a mobile phone, representing the victim in an unfavorable light, discrediting her honor and dignity, in this regard, the appearance of "specialized " sites on which, under the slogan "take revenge on your ex / your ex", it is proposed to post intimate photos of a once close person should be particularly noted.;
- posting on photo and video hosting sites, as well as on social networks, photos and videos with the victim's participation made by a cyberbuller (openly or secretly) and representing the victim in an unsightly way, for example, drunk, while changing clothes

in the locker room, in the shower, in the toilet, etc.; separately, it should be said about videos and photos showing rape, beating, bullying or humiliation of the victim (this problem is especially relevant for the teenage environment).

2.5. Sexting and grooming

The word "sexting" (from the English sex and texting) means communication on the topic of sex via a mobile phone or the Internet. Almost a third of Russian schoolchildren have met or received personally sexual messages on the Internet, and more than 15% - once a month and more often. 4% of children send or write sexual messages themselves. In terms of the percentage of children receiving or encountering sexual messages on the Internet, Russia is ahead of all European countries [23]. Recently, this threat is especially relevant for children of primary and secondary school age, who are actively exploring the expanses of the Internet, but at the same time are not ready to respond adequately to such messages, i.e. online children face those threats from which they are practically protected in real life.

3. RELATED WORKS

The task of detecting malicious information on the Internet can be reduced to the classification of web pages, in which a number of categories are pre-determined by the system administrator as containing illegitimate content. In this field, there are many works devoted to the construction of both expert systems and fully automatic systems.

The CONSTRUE system presented in [24] is based on production rules created manually by an expert operator. This system is designed to classify economic and financial news and correlate the analyzed text to one of 674 categories. The classification accuracy for the CONSTRUE system is more than 90%. The disadvantage of such a system is that its maintenance in a consistent state requires the regular involvement of specialists who perform the addition and correction of production rules.

The approach to content categorization with automatic generation of classification rules is considered by researchers C. Apté, F. Damerau and S.M. Weiss [25]. Their proposed rule format is disjunctive normal form (DNF). The algorithm for forming rules is based on the sequential replacement of one of the conjuncts and the further addition of a new conjunct until one hundred percent coverage of the training sample is built (i.e., such a set of rules that will ensure an error-free classification of training elements). This algorithm performs a heuristic search for such rules: the algorithm does not provide finding the minimum number of DNF conjuncts. In addition, unlike a decision tree, conjuncts united by a single rule using this algorithm are not mutually exclusive.

Predicates reflecting the signs were used as elementary conjuncts (atoms):

1) occurrences of a certain word (or phrase) from a local dictionary (a set of words containing concepts specific to one category) in the analyzed text;

2) exceeding the frequency of occurrence of a certain expression within the analyzed text by the specified threshold value.

The proposed approach allows you to save the presentation of the rules in a format convenient for analysis by experts. At the same time, with the described method of generating rules, the generalizing ability of the system and the ability to process noisy data are lost.

The authors of the article [26] propose to accept the analyzed document as an array of real-valued coefficients, which represent the relative and absolute frequencies of occurrence of certain words in the classified text. Among such coefficients were highlighted:

- frequency of the word (TF, from the English Term Frequency);
- inverse document frequency (IDF, from the English Inverse Document Frequency);
- the importance of the word (TD, from the English Term Discrimination), where TD = TF × IDF; and some others.

In [27], an approach is outlined that allows attributing to each word its integral weight, including the probability of the occurrence of this word, both within a certain category and within the entire collection of documents, and taking into account other categories.

The training, namely the Bayesian classifier and the decision tree, within the framework of the text categorization task was performed in [27]. The authors of this article emphasize that the decision tree demonstrates the best performance on large sets of training data, and the Bayesian classifier demonstrates the best performance on smaller data sets. Moreover, for the Bayesian classifier, with an increase in the number of processed features, a situation of retraining is observed (on the control set, the performance of the classifier decreases), and for the decision tree, under the same conditions

and a sufficient amount of training sample, there is an increase in classification efficiency indicators.

The applicability of another popular machine learning method, namely the support vector machine, to the problem of text classification is investigated in [28], where the author highlights the ability of SVM to learn on both high-dimensional and sparse feature vectors. The solution to the problem of text classification in most cases has the form of linearly separable areas, for which the SVM can be used.

The article by [28] describes two types of convolution neural networks: direct signal propagation and with the transformation of a "bag" of words on the convolution layer. As a result of experiments, the authors revealed that the first type of neural network demonstrates greater performance in terms of classification indicators compared to the second type of neural network.

[29] presents a method for extracting features within the framework of the text categorization task. The proposed modification of the genetic algorithm, as experiments show, makes it possible to achieve a more compact representation of training vectors in terms of their dimension and improve the quality of classification of the analyzed text.

A common limitation for the above-mentioned works devoted to the application of machine learning methods to the problem being solved is the use of single-component classifiers, which makes it impossible to train the model in parts and, in turn, makes it difficult to parallelize this process.

The analysis of works in this subject area shows the relevance of the topic under consideration. At the same time, despite their diversity, the task of developing a methodology designed to detect malicious IO on the Internet and combining machine learning methods and their combination remains a high priority in the research community.

4. ANALYSIS OF EXISTING APPROACHES

To solve the problem of automatic identification of destructive information on the example of searching and identifying profanity, the existing search methods used to find specific words in the text were analyzed [30-31]. It is revealed that these methods are broadly divided:

− methods based entirely on expert information processing, the result of which is the creation of "black lists", registries, etc.;

− automated methods of interest in our work, including the so-called thematic search (by dictionary) and intelligent data processing methods.

Dictionary search, which consists in the system's search for an exact match of a word from the dictionary and its identification in the text, is one of the most common types of detection of destructive information [32-34]. However, to solve the problem of identifying destructive content on the example of profanity, the use of thematic search in a "pure" form is impractical due to the following reasons:

- the need for constant replenishment of the dictionary due to the emergence of new forms and methods of word formation (at the moment there are more than 250 basic profanity words forming various combinations and word forms);

- a high probability of omitting a profanity due to the use of a new form of it in the text under study, which is not known to the system.

Systems based on intelligent data processing methods have the following advantages [35]:

- symbolic (semantic) processing of information in a form close to human thinking;

- developed communication skills that allow for an intensive dialogue with users, during which the knowledge available and acquired by the system is clarified;

- forming requests to the system and receiving answers (problem solutions) in a natural language close to human communication;

- the ability to self-study, i.e. to automatically replenish and acquire new knowledge based on the accumulated experience of analyzing and solving user tasks by the system;

- the ability to adapt (adaptability) of the system to objective changes in the subject (problem) area of the functioning of the system, etc.

It is obvious that an effective automated system for identifying profanity should support the listed functionality, in particular:

- automatic replenishment of the dictionary in case of detection of a "dangerous" word;

- request from an expert to check for the correctness of the replenishment of the dictionary (if necessary);

- providing detailed reports on the results of analytical search.

Thus, to solve this problem, an intelligent algorithm for identifying profanity in the text was developed, implementing a modified dictionary search.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5. DATA MINING AND COLLECTION

5.1. Dataset

In the first part of our research, we collected data with the goal of creating a dataset for feeding machine learning methods. The data were collected from social networking sites and classified into two classes as the data that contains destructive idea, and the data that does not contain destructive idea.

Figure 3 illustrates data collection process of our exploration that divided into data source identification, parsing, and dataset development. In the first stage, we define data source that we collect for dataset creation. As we said before, in our case social networking sites were data source. After defining data source, we download data by using a parser. After getting the necessary data, we classify them into two categories. necessary data for feeding machine learning

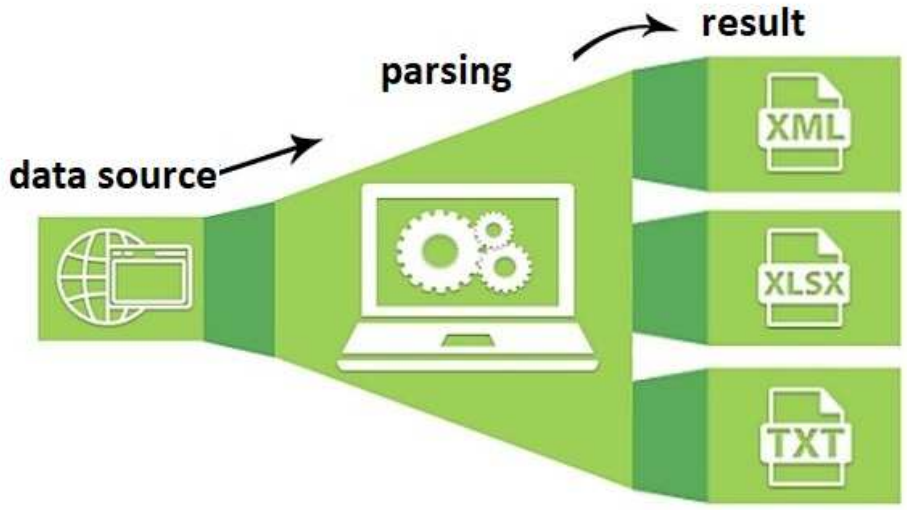

Figure 3: Data Collection

Figure 4 demonstrates data collection flowchart. In order to get data, we connect to the social networking sites via corresponding APIs. We download the data in .json format using the API. After downloading we classify the data by hand into two categories. In the end, after the data classification, we can start to create a dataset in convenient form for ourselves. For collecting data from Vkontakte social network, we used VK API that allow to get 1% of all the data for research purposes.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

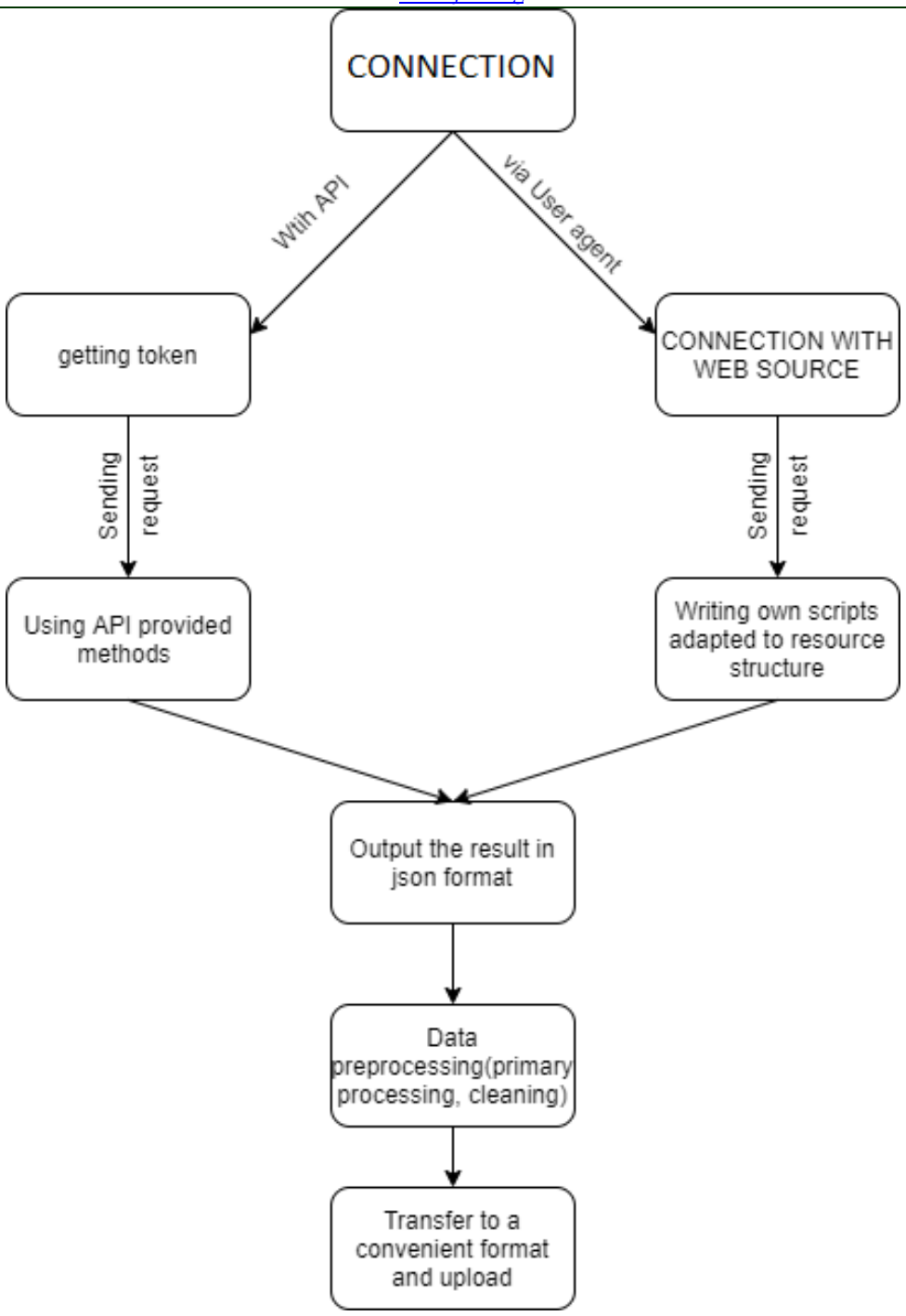

Figure 4: Data Collection Flowchart

For collecting data from Instagram, we used another method as Instagram restrict downloading data using API. In the next sections, we explain how we get data from Instagram.

Table I shows rule base that we use to do request to get data. There, we explain methods that used and give their descriptions. There, we use different methods as users.get, account.getinfo, get, copy_history_depth, wall.getComment for downloading posts and comments of the Instagram users.

Table II demonstrates URL addresses and their actions that applied to download the necessary data.

Table 1: Rule Base

Methods	Description
users.get	Method for using the user details.
account.getinfo	There exists a system for retrieving current details on the owner.
getMembers	Returns a registry of community organization participants.
get	The performance is any sort of collection of public posts or group wall posts.
copy_history_depth	The size of the copy history array returned if the record is a repost of a record from another wall is determined by the value of the copy history bias array.
wall.getComment	This software gets details regarding messages on the panel.

In addition, the app will utilize a specifically developed URL scheme to show a snap if special authorization is granted.

URL	Action
media?id	Plays the unique ID of an Instagram message, opens up the Instagram software, and delivers.
user?username	It will open the Instagram app and load up the user with the defined value of @username
location?id	Upon opened, the app opens to the geo's (location) feed, which contains the defined ID.
tag?name	This action would open the Instagram app, load a page with the 'tag' for the stated name, and launch a page with the hashtag for that name.

Next listing presents sample of query.

```
curl -X GET \
"https://graph.facebook.com/v9.0/instagram_oembed?url=
https://www.instagram.com/p/fA9uwTtkSN/&access_token= IGQVJ..."
```

Next code demonstrates sample of a response.

```
{ "version": "1.0", "author_name": "diegoquinteiro", "provider_name": "Instagram", "provider_url": "https://www.instagram.com/", "type": "rich", "width": 658, "html": "<blockquote class=\"instagram-media\" data-instgrm-ca...", "thumbnail_width": 640, "thumbnail_height": 640 }
```

Therefore that we already have received information via the use of an online platform, the idea of accessing should be the same for users of that raw material. The request library is intended that will be used to establish a connection with a website and to begin doing business with it.

```
import requests
user_id = 12345
url =
'http://www.kinopoisk.ru/user/%d/votes/list/ord/date/ page/2/#list' % (user_id) #
        headers = { 'User-Agent': 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10.9; rv:45.0) Gecko/20100101 Firefox/45.0' }
        r = requests.get(url, headers = headers)
```

Using this document, we were able to get all of the headers for this page's HTML code. In Listing 1 we demonstrate source of code to get data in Instagram.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

```
from lxml import html
test = '''
    <html>
        <body>
            <div class="first_level">
                <h2 align='center'>one</h2>
                <h2 align='left'>two</h2>
            </div>
            <h2>another tag</h2>
        </body>
    </html>
'''
tree = html.fromstring(test)
tree.xpath('//h2') # все h2 теги
tree.xpath('//h2[@align]') # h2 теги с атрибутом align
tree.xpath('//h2[@align="center"]') # h2 теги с атрибутом align равным "center"

div_node = tree.xpath('//div')[0] # div тег
div_node.xpath('.//h2') # все h2 теги, которые являются дочерними div нода
```

Listing 1. Collecting headers.

Next part of code demonstrates save the data in .txt file.

```
f = open("demofile2.txt", "a") # opening the file for editing
f.write("---") #record data
f.close() # close
```

Next section of the research explains data preprocessing part.

5.2. Data Preprocessing

Figure 5 illustrates word2vec model of our collected dataset that contains two categories the content that contains destructive ideas and the content that contains neutral texts.

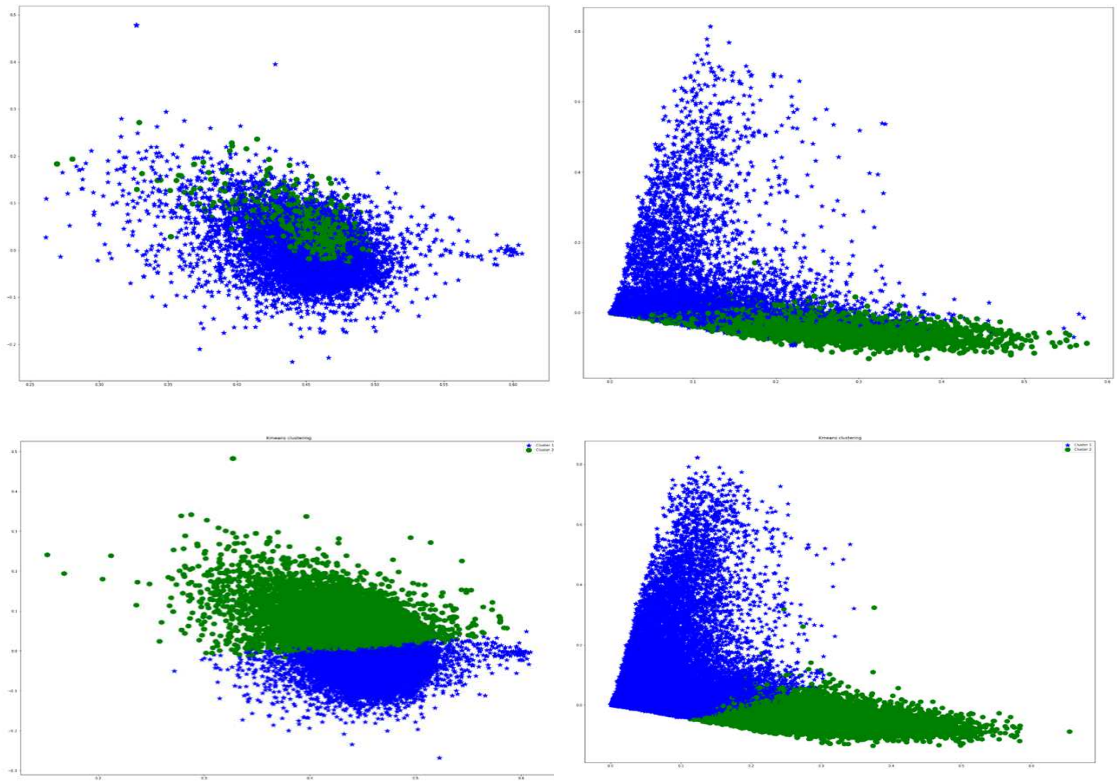

Figure 5: Graphical Representation Of Word2vec Vectors For The Collected Dataset

Figure 6 demonstrates representation of the collected texts. Blue part means neutral texts, red parts the texts that contain destructive ideas. In our case, it is imbalanced dataset, as it is difficult to find destructive texts in big volume.

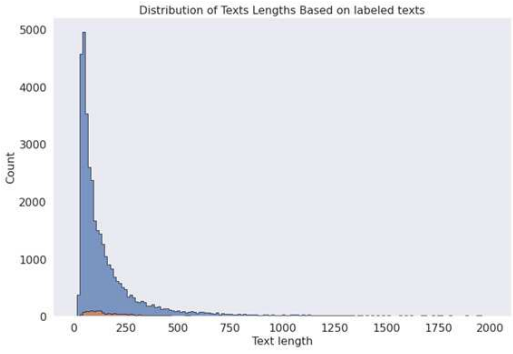

Figure 6: Representation Of Collected Data By Categories

Next two figures (Figure 7 and Figure 8) demonstrates top unigrams and bigrams in the collected dataset.

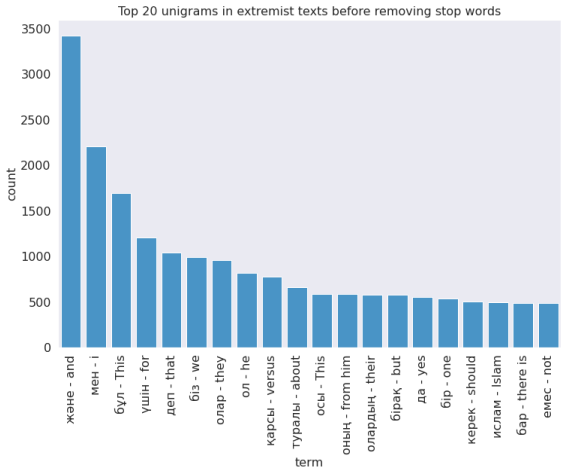

Figure 7: Top Unigrams

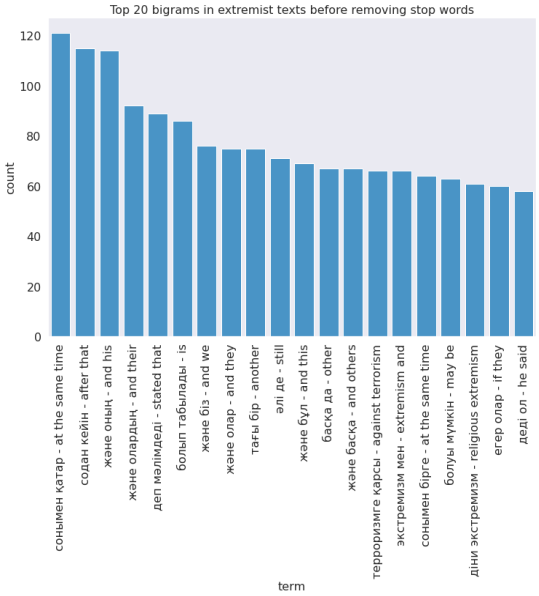

Figure 8: Top Bigrams

6. EXPERIMENT RESULTS

As we are going to classify texts into two categories Equation (1) can express problem statement of our task.

$$Y_i = F(x_i) \tag{1}$$

When, $Y_i = 1$ is destructive contents, $Y_i = 0$ neutral content.

The problem will be solved by using supervised learning methods. In our research we used six supervised learning methods as SVM, Decision tree, Random forest, KNN, Naïve Bayes, and Logistic regression.

6.2 Evaluation metrics

Equations (2) to (5) demonstrates metrics for assess the proposed solution. As an evaluation criteria we used precision, recall, F1-score, and accuracy.

$$precision = \frac{TP}{TP + FP} \tag{2}$$

$$recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = \frac{2\,precision \cdot recall}{precision + recall} \quad (4)$$

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad (5)$$

Here, TP – true positives, TN – true negatives, FN – false negatives, FP – false negatives. .

6.2 Classification Results

This section demonstrates the classification results and compare the machine learning algorithms for the given problem. Table II demonstrates the results of classification for each machine learning method.

Table 2: Results Of Destructive Content Detection

ML Method	*Accuracy*	*Precision*	*Recall*	*F1 score*
SVM	0.7815	0.7457	0.6954	0.4018
Decision Tree	0.8064	0.6942	0.6927	0.8047
Random Forest	0.6874	0.7062	0.4462	0.6815
KNN	0.8234	0.7021	0.7108	0.4213
Naïve Bayes	0.7756	0.6871	0.7312	0.7218
Logistic Regression	0.7964	0.7045	0.7328	0.6824

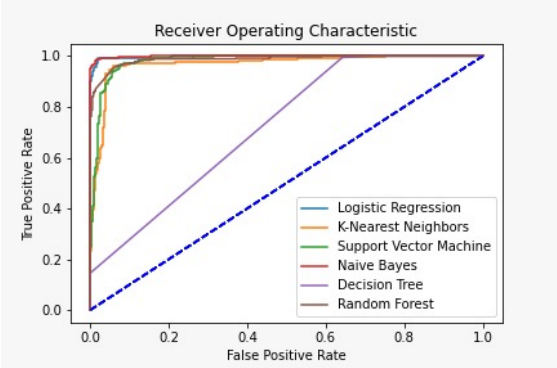

Figure 7: ROC Curves Result

Figure 7 compares AUC-ROC curves of each method. As it is illustrated in the figure KNN and Logistic regression demonstrates the best AUC-ROC comparing the other methods. Thus, the results demonstrates that automated identification of destructive contents in the internet by using machine learning methods is solvable problem. In further we are going to apply deep learning techniques to improve the obtained results.

7. CONCLUSION

Due to the current trend of integration of social networks and cloud services, it can be assumed that in the near future there will be proposals for some kind of universal platform or environment for working on the Internet, combining a social network, email, communication system (messaging, voice and video communication), search engine, remote data storage, personal assistant, financial management system, including bank accounts and cards, as well as other services. This perspective determines a significantly higher level of requirements for ensuring user security when working on the Internet. In this research, we apply data mining techniques to collect data, and create dataset to feed machine learning models, and apply the machine learning models to detect destructive contents. The results show that, the given problem is solvable and machine learning techniques are useful to detect destructive contents on the internet. Maximum accuracy 82.34% that is useful for the given problem. In the next part of our research, we are going to use deep learning techniques to get more accuracy and increase the detection process.

REFERENCES:

[1] Yu, X., Wang, J., Wen, S., Yang, J., & Zhang, F. (2019). A deep learning based feature extraction method on hyperspectral images for nondestructive prediction of TVB-N content in Pacific white shrimp (Litopenaeus vannamei). Biosystems Engineering, 178, 244-255.

[2] Yu, M., Huang, Q., Qin, H., Scheele, C., & Yang, C. (2019). Deep learning for real-time social media text classification for situation awareness–using Hurricanes Sandy, Harvey, and Irma as case studies. International Journal of Digital Earth, 12(11), 1230-1247.

[3] Khan, J. Y., Khondaker, M., Islam, T., Iqbal, A., & Afroz, S. (2019). A benchmark study on machine learning methods for fake news detection. arXiv preprint arXiv:1905.04749.

[4] Salloum, S. A., Alshurideh, M., Elnagar, A., & Shaalan, K. (2020, March). Machine Learning and Deep Learning Techniques for Cybersecurity: A Review. In AICV (pp. 50-57).

[5] Omarov, B., Saparkhojayev, N., Shekerbekova, S., Akhmetova, O., Sakypbekova, M., Kamalova, G., ... & Akanova, Z. (2022). Artificial Intelligence in Medicine: Real Time

Electronic Stethoscope for Heart Diseases Detection. CMC-COMPUTERS MATERIALS & CONTINUA, 70(2), 2815-2833.

[6] Khan, W., Ahmed, A. A. A., Vadlamudi, S., Paruchuri, H., & Ganapathy, A. (2021). Machine Moderators in Content Management System Details: Essentials for IoT Entrepreneurs. Academy of Entrepreneurship Journal, 27(3), 1-11.

[7] Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2020). Detection of suicide ideation in social media forums using deep learning. Algorithms, 13(1), 7.

[8] Mohan, A., Singh, A. K., Kumar, B., & Dwivedi, R. (2021). Review on remote sensing methods for landslide detection using machine and deep learning. Transactions on Emerging Telecommunications Technologies, 32(7), e3998.

[9] Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential biases in machine learning algorithms using electronic health record data. JAMA internal medicine, 178(11), 1544-1547.

[10] Govil, K., Welch, M. L., Ball, J. T., & Pennypacker, C. R. (2020). Preliminary results from a wildfire detection system using deep learning on remote camera images. Remote Sensing, 12(1), 166.

[11] Subramani, S., Michalska, S., Wang, H., Du, J., Zhang, Y., & Shakeel, H. (2019). Deep learning for multi-class identification from domestic violence online posts. IEEE Access, 7, 46210-46224.

[12] Omarov, B., Altayeva, A., Demeuov, A., Tastanov, A., Kassymbekov, Z., & Koishybayev, A. (2020, December). Fuzzy Controller for Indoor Air Quality Control: A Sport Complex Case Study. In International Conference on Advanced Informatics for Computing Research (pp. 53-61). Springer, Singapore.

[13] Mussiraliyeva, S., Omarov, B., Yoo, P., & Bolatbek, M. (2021). Applying machine learning techniques for religious extremism detection on online user contents. Computers, Materials and Continua, 70(1), 915-934.

[14] Amanpreet Singh, Maninder Kaur. Content-based Cybercrime Detection: A Concise Review. International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-8 June, 2019

[15] Ofli, F., Alam, F., & Imran, M. (2020). Analysis of social media data using multimodal deep learning for disaster response. arXiv preprint arXiv:2004.11838.

[16] Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., ... & Wang, C. (2018). Machine learning and deep learning methods for cybersecurity. Ieee access, 6, 35365-35381.

[17] Omarov, B., Omarov, B., Issayev, A., Anarbayev, A., Akhmetov, B., Yessirkepov, Z., & Sabdenbekov, Y. (2020, November). Ensuring comfort microclimate for sportsmen in sport halls: comfort temperature case study. In International Conference on Computational Collective Intelligence (pp. 626-637). Springer, Cham.

[18] Wang, Y., Yang, W., Ma, F., Xu, J., Zhong, B., Deng, Q., & Gao, J. (2020, April). Weak supervision for fake news detection via reinforcement learning. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 01, pp. 516-523).

[19] Nurtas, M., Baishemirov, Z., Aizhan, Y., & Aizhan, A. (2020, September). 2-D Finite Element method using" eScript" for acoustic wave propagation. In Proceedings of the 6th International Conference on Engineering & MIS 2020 (pp. 1-7).

[20] Hussain, B., Du, Q., Sun, B., & Han, Z. (2020). Deep learning-based DDoS-attack detection for cyber–physical system over 5G network. IEEE Transactions on Industrial Informatics, 17(2), 860-870.

[21] Murzamadieva, M., Ivashov, A., Omarov, B., Omarov, B., Kendzhayeva, B., & Abdrakhmanov, R. (2021, January). Development of a system for ensuring humidity in sport complexes. In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 530-535). IEEE..

[22] Ren, K., Zheng, T., Qin, Z., & Liu, X. (2020). Adversarial attacks and defenses in deep learning. Engineering, 6(3), 346-360.

[23] Otoum, Y., Liu, D., & Nayak, A. (2019). Toktarova, A., Beissenova, G., Kozhabekova, P., Makhanova, Z., Tulegenova, B., Rakhymbek, N., ... & Azhibekova, Z. (2021). Automatic offensive language detection in online user generated contents. Journal of Theoretical and Applied Information Technology, 99(9), 2054-2067.

[24] MacAvaney, S., Yao, H. R., Yang, E., Russell, K., Goharian, N., & Frieder, O. (2019). Hate

speech detection: Challenges and solutions. PloS one, 14(8), e0221152.
[25] Singh, J. P., Kumar, A., Rana, N. P., & Dwivedi, Y. K. (2020). Attention-based LSTM network for rumor veracity estimation of tweets. Information Systems Frontiers, 1-16.
[26] Canhoto, A. I., & Clear, F. (2020). Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. Business Horizons, 63(2), 183-193.
[27] Bekbauov, B., Berdyshev, A., & Baishemirov, Z. (2016). Numerical Simulation of Chemical Enhanced Oil Recovery Processes. In DOOR (Supplement) (pp. 664-676).
[28] Yuan, S., & Wu, X. (2021). Deep learning for insider threat detection: Review, challenges and opportunities. Computers & Security, 102221.
[29] NURTAS, M., IROV, Z. B., ALPAR, S., & TOKMUKHAMEDOVA, F. (2021). Numerical simulation of wave propagation in mixed porous media using finite element method. Journal of Theoretical and Applied Information Technology, 99(16), 4163-4172.
[30] Mühlhoff, R. (2020). Human-aided artificial intelligence: Or, how to run large computations in human brains? Toward a media sociology of machine learning. new media & society, 22(10), 1868-1884.
[31] Neupane, D., & Seok, J. (2020). Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review. IEEE Access, 8, 93155-93178.
[32] Kumar, A., Singh, J. P., Dwivedi, Y. K., & Rana, N. P. (2020). A deep multi-modal neural network for informative Twitter content classification during emergencies. Annals of Operations Research, 1-32.
[33] Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. Journal of Big data, 7(1), 1-29.
[34] Omarov, B., Omarov, B., Shekerbekova, S., Gusmanova, F., Oshanova, N., Sarbasova, A., ... & Sultan, D. (2019, October). Applying face recognition in video surveillance security systems. In International Conference on Objects, Components, Models and Patterns (pp. 271-280). Springer, Cham.
[35] Naeem, B., Khan, A., Beg, M. O., & Mujtaba, H. (2020). A deep learning framework for clickbait detection on social area network using natural language cues. Journal of Computational Social Science, 1-13.