DOI: https://doi.org/10.5281/zenodo.12223614

EMAIL FILTERING METHODS AND THEIR COMPARISON ANALYSIS

Istamov Mirjahan Muminjan
Tohirov Quvonchbek Musurmon o‘g‘li
Sultonov Hayotjon Baxodir o‘g ‘li
Students at the Tashkent University of Information Technologies named after
Mukhammad al-Kharezmy

Abstract: *This prevents email attacks from occurring. Each e-mail message is checked by a program created by a fully machine learning algorithm. In the development of this program, I used the pandas and sklearn libraries, which are part of the machine learning libraries.*

Key words: *e-mail filters, Spam, networks, SMTP*.

Nowadays, as information and communication systems develop, the demand for using electronic information is increasing and almost all countries around the world support the electronic mail exchange system and the flow of electronic mail between organizations is increasing day by day. It is the only e-mail exchange system recognized by all countries in the world is email. There are other types of e-mail exchange systems, but other systems are not recognized by all countries.

E-mail is a technology that allows for the real-time online exchange of e-mails across computer networks, that is, the sending and receiving of e-mails. Modern e-mails are over the Internet or works through computer networks. Anyone who can work on a computer can open a free e-mail box and use it freely. For this, it is sufficient to register from one of the portals providing e-mail service. After registration, e-mail can be accessed from any computer connected to the Internet. There are many such portals. The most famous of them are:

- http://mail.yandex.ru
- http://www.umail.uz
- http://mail.rambler.ru
- http://mail.google.com
- http://mail.yahoo.com
- http://www.mail.ru

The advantages of e-mail are that it is easy to remember the names of addresses and that you can send files in any format. Fast and reliable delivery of messages, ease of use and similar features can be cited. Disadvantages of e-mail are the mass distribution of advertisements and viruses, the limitation of the size of the message, and the possibility of using it for malicious purposes if a stranger enters the mailbox.

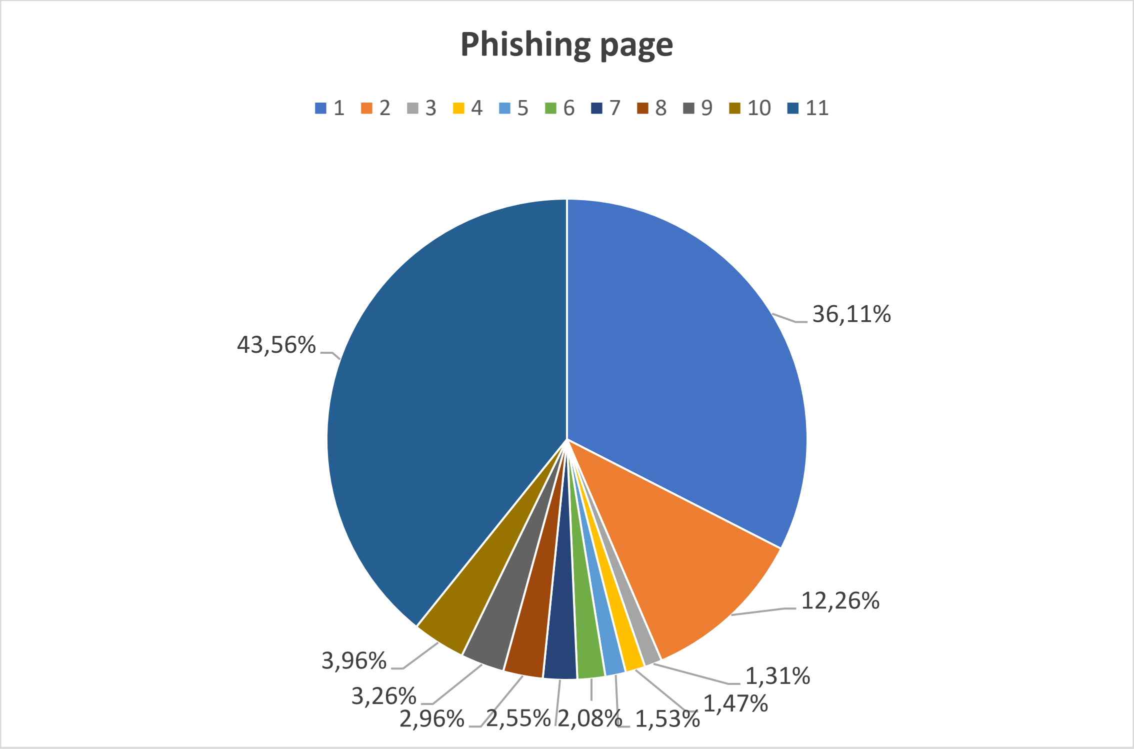

Fig 1.1. The top level with more spam and phishing pages domains

Over the past decade, email spam has become the norm is a serious problem for mail users . Every day too many spam messages are sent to users' e-mails. Most spam is sent to closed email addresses. Spam makes up 78% of all sent messages. In the first half of 2010, 88-92% of all electronic messages were sent, according to the Task Force on the Abuse of Messages.

As of 2020, spam accounted for 50.37% of email traffic, down 6.14% from 2019. Most of the spam, 21.27%, came from Russia. Kaspersky Lab detected 184, 435, 643 malicious links. Mail antiviruses often detected messages containing malware from the Trojan.Win32.Agentb family. Antiphishing system blocked 434898635 attempts to go to fake resources. 18.12% of Internet stores are often subject to phishing attacks. This is why email filtering is essential.

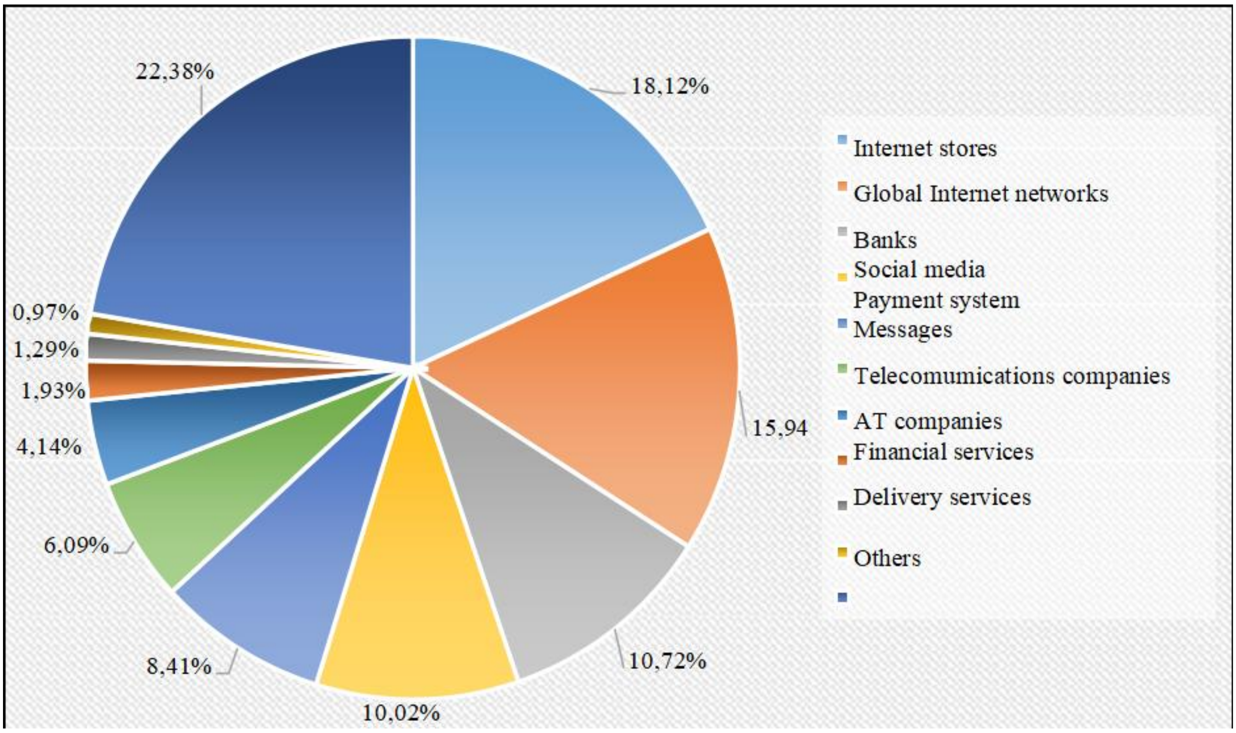

Fig 1.2. Distribution of organizations attacked by spammers

Spam is the systematic transmission of commercial electronic messages by persons who hide or falsify their real coordinates. In a broader sense, it is an electronic document sent to an email address in the form of an advertisement from an unknown person or organization. There may also be spam messages within the local network. The bad thing about spam is that it floods users' mailboxes with unwanted messages. This, in turn, leads to spending the purchased traffic on opening unnecessary messages. If the number of spam messages is too high and the interval between incoming spam messages is too short, it will cause the mail server to crash and the Internet channel to overload. Spam messages have the following types:

- e-mail spam;
- spam in social networks;
- spam on forums;
- spam distributed through comments on sites;
- spam in the form of catalogs and newsletters;
- sms spam;

People or organizations that send spam are called spammers. It is possible to increase the counter that counts the visitors to the social page with the help of spam, and it is possible to see income through this. That's why anti-spam filters are installed when creating a site. An example of this is the introduction of the CAPTCHA system (writing numbers or letters that the machine does not understand) when writing comments. In this way, it will be possible to control whether the person leaving the message is a human or a robot. Basically, spam is sent to users through a script, an algorithm is created and presented to the public.

Email spam is a type of email spam that consists of messages sent to a large number of recipients via email. Clicking on a link in a spam message may redirect the user to a phishing website or a site that hosts malware. Spam in e-mail can contain malware in the form of embedded scripts or other executable files. A spam detection email usually contains a large number of links. Most spam filtering approaches currently in use focus on the classification method, where text messages are the primary content. Classified spam messages may consist of neutral words.

Spam not only harms the majority of e-mail users, but also creates a load on the organization's information system infrastructure and causes loss of information system efficiency.

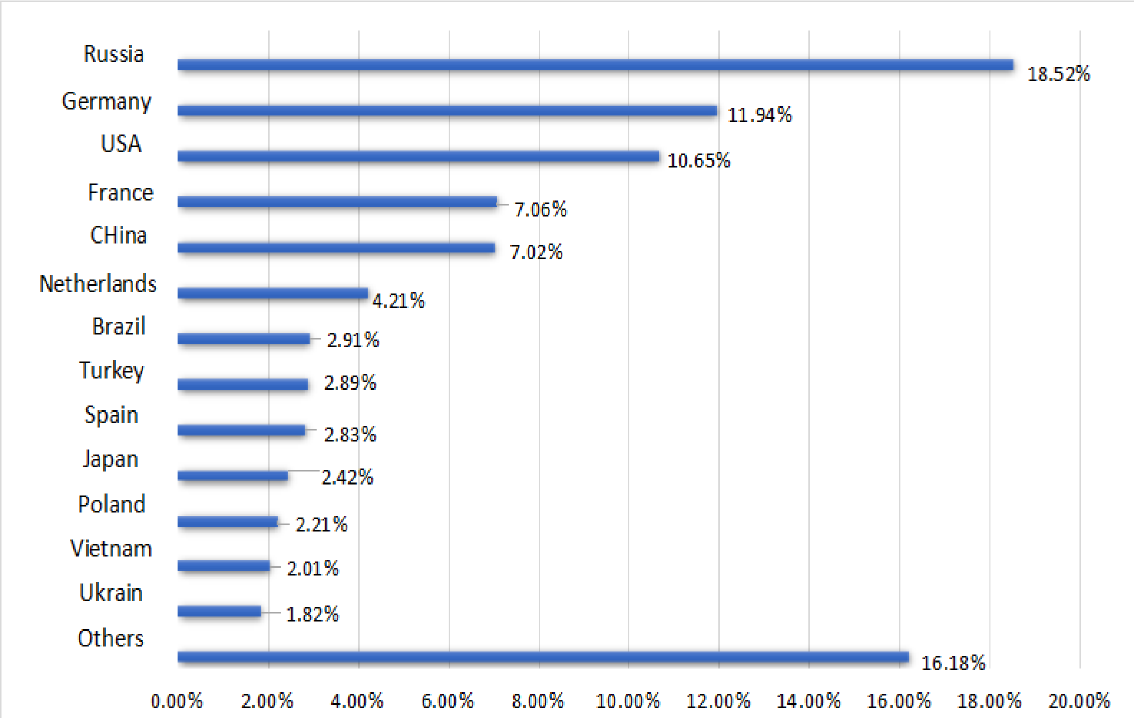

Fig 1.3. Share of spam messages distributed worldwide

E-mail filtering is a real-time analysis of incoming e-mail messages according to the specified criteria and the result of the analysis is the process of accepting or rejecting an electronic message. It is mostly used in the process of automatic processing of incoming e-mail messages, but filtering is used in addition to the method of combating spam messages to add a human factor and to e-mail messages and outgoing messages.

E-mail filters are usually used to detect and eliminate viruses or spam messages from incoming e-mail messages. In some organizations, as a result of employees' non-compliance with established regulatory legal documents and laws, spam messages spread to the organization's local or corporate network. Users are encouraged to use email filters to sort messages into folders based on subject or other criteria. Mail filters are installed by the user as a separate program or as part of a mail program. In e-mail programs, users can easily create programs that automatically filter mail according to selected criteria. Most email programs have automatic spam filtering. ISPs can also install mail filters on mail agents as a service to their customers. Due to the growing threat of fraudulent websites, ISPs filter URLs in email messages to prevent threats from malicious users. Mail filters can work with incoming and outgoing traffic. Includes a system of filtering incoming e-mail messages or the process of scanning messages from the Internet to users who are protected from legal interception. Outbound email filtering involves rescanning email messages from local users to prevent potentially dangerous messages from being delivered to other users on the Internet. One method of filtering outgoing e-mail messages commonly used by ISPs is transparent SMTP (Simple Mail Transfer Protocol) proxying, in which e-mail traffic is intercepted and filtered through transparent proxy servers on the network. Enterprises are more themselves uses e-mail filtering software for the protection of employees and information technology.

In recent times, unwanted commercial bulk emails called spam has become a huge problem on the internet. The person sending the spam messages is referred to as the spammer. Such a person gathers email addresses from different websites, chatrooms, and viruses. Spam prevents the user from making full and good use of time, storage capacity and network bandwidth. The huge volume of spam mails flowing through the computer networks have destructive effects on the memory space of email servers, communication bandwidth, CPU power and user time. The menace of spam email is on the increase on yearly basis and is responsible for over 77% of the whole global email traffic. Users who receive spam emails that they did not request find it very irritating. It is also resulted to untold financial loss to many users who have fallen victim of internet scams and other fraudulent practices of spammers who send emails pretending to be from reputable companies with the intention to persuade individuals to disclose sensitive personal information like passwords, Bank Verification Number (BVN) and credit card numbers.

CONCLUSION

Detecting spam in emails using machine learning represents a significant advancement in enhancing cyber security and improving user experience. Traditional spam detection methods, such as rule-based filtering, have limitations in adapting to the evolving tactics of spammers. Machine learning approaches address these limitations by leveraging data-driven techniques to identify and filter spam with greater accuracy and efficiency.

Machine learning models, such as Naive Bayes, Support Vector Machines, Decision Trees, Random Forests, and neural networks, analyze vast amounts of email data to recognize patterns and characteristics typical of spam. These models are trained on labeled datasets containing examples of both spam and non-spam emails, allowing them to learn and generalize patterns that differentiate the two categories. By continuously updating and refining these models with new data, machine learning systems can adapt to new spam techniques and maintain high detection rates.

REFERENCES

1. "Machine Learning for Email: Spam Filtering and Priority Inbox" by Trevor Grant, Simon Walk, and Luis Vargas
2. https://www.hindawi.com/journals/scn/2022/1862888/
3. "Data Mining: Practical Machine Learning Tools and Techniques" by Ian H. Witten, Eibe Frank, and Mark A. Hall
4. "Learning from Data: A Short Course" by Yaser S. Abu-Mostafa, Malik Magdon-Ismail, and Hsuan-Tien Lin
5. "Neural Networks and Deep Learning: A Textbook" by Charu C. Aggarwal
6. "Machine Learning: An Algorithmic Perspective" by Stephen Marsland
7. https://www.sciencedirect.com/science/article/pii/S2405844018353404
8. https://researchgate.net/publication/348576063_Comparative_Analysis_of_Detection_of_Email_Spam_With_the_Aid_of_Machine_Learning_Approaches_Comparative_Analysis_of_Detection_of_Email_Spam_With_the_Aid_of_Machine_Learning_Approaches
9. Email Spam Detection Using MachineLearning Mrs. Anitha Reddy1*, Kanthala Harivardhan Reddy2 , A. Abhishek3 , Myana Manish4 , G. Viswa Sai Dattu5 , Noor Mohammad Ansari6