**Phishing email detection**

1. **Descrierea ideii de proiect**

Emailurile de tip phishing[[1]](#footnote-1) sunt mesaje frauduloase create pentru a induce în eroare utilizatorii și a obține informații sensibile, cum ar fi parole, informații bancare sau date personale. Aceste atacuri sunt tot mai sofisticate și se bazează pe tehnici de inginerie socială, iar detectarea automată a acestor mesaje a devenit o provocare esențială în securitatea cibernetică.

Soluțiile de detectare a phishingului sunt de obicei integrate în sistemele de email și securitate cibernetică și urmăresc identificarea și blocarea emailurilor suspecte înainte de a ajunge în inboxurile utilizatorilor.

Proiectul își propune să construiască un sistem capabil să identifice emailurile de phishing folosind tehnici avansate de învățare automată și procesare a limbajului natural (NLP - Natural Language Processing). Această abordare este esențială pentru că poate permite detectarea automată și rapidă a acestor mesaje, reducând riscul pentru utilizatori și organizații.

1. **State-of-the-art**

În literatura de specialitate[[2]](#footnote-2) abordările clasice pentru detectarea phishingului se împart în două categorii: *blacklists* și tehnici *siganture-based*. Blacklisting-ul presupune întocmirea unei liste cu resurse suspicioase folosite în atacuri de phishing anterioare. Conținutul nou suspicios poate fi comparat cu blacklist-a pentru a se confirma validitatea acestuia. Din păcate, managerierea blacklist-urilor a devenit foarte dificilă odată cu rapiditatea ștergerii vechilor și a creării de noi adrese suspicioase. Pe de altă parte, abordarea signature-based se concentrează pe folosirea caracteristicilor asociate cu phishing-ul, precum adrese de email, URL-uri, link-uri, pagini web, în combinație cu regulile de detectare a atacurilor de phishing. Deși această abordare este mai eficientă decât folosirea blacklisting-ului, a dat dovadă de o rată mare de rezultate fals-pozitive.

Abordările tradiționale de detectare a phishingului au scăzut în popularitate, deoarece se bazau pe un efort uman foarte mare, în favoarea tehnicilor moderne de machine learning și deep-learning. Algoritmii de ML pot fi folosiți în antrenarea modelelor care detectează automat emailurile de phishing, prin învățarea caracteristicilor și a pattern-urilor din seturi de date pe care aceștia le analizează. Mai mult decât atât, DL este capabil să extragă automat cele mai importante caracteristici din date neprelucrate. În prezent, deep neural networks sunt folosite în diverse domenii datorită performanței înalte. Se cercetează rezultatele promițătoare[[3]](#footnote-3) ale modelelor CNN, LSTM și GRU.

Cele mai cunoscute produse care protejează email-urile de atacuri cibernetice sunt **Microsoft Defender[[4]](#footnote-4)** și **Proofpoint[[5]](#footnote-5).**

1. **Setul de date**

Unul dintre cele mai populare seturi de date de pe web este Phishing Email Dataset[[6]](#footnote-6), dar acesta etichtează ca *phishing* și email-urile care sunt *spam*. Pentru proiectul nostru, avem în vedere specific detectarea email-urilor care pot reprezenta atacuri de phishing. În timp ce spamul este pur și simplu nedorit, phishing-ul este creat în mod deliberat de un actor rău intenționat pentru a afecta o companie sau un individ prin obținerea de informații sensibile.[[7]](#footnote-7)

Abordarea pe care am ales-o a fost construirea unui set de date complex folosind combinarea a 4 seturi populare, disponibile online, pentru a evita overfitting-ul. Aceste seturi de date sunt:

* **Nazario phishing corpus[[8]](#footnote-8)**
* **Clair fraudulent email corpus[[9]](#footnote-9)**
* **Spamassassin corpus[[10]](#footnote-10)**
* **Enron email dataset[[11]](#footnote-11)**

Vom eticheta datele ca „ham” și „phishing” și le vom împărți într-un set de date de antrenament și unul de test (în proporție de 8:2). Ulterior vom analiza mai multe abordări pentru detectarea email-urilor de phishing asupra setului de date obținut.

1. <https://www.techtarget.com/searchsecurity/definition/phishing> [↑](#footnote-ref-1)
2. <https://www.mdpi.com/2079-9292/12/21/4545> [↑](#footnote-ref-2)
3. <https://www.mdpi.com/2079-9292/12/20/4261> [↑](#footnote-ref-3)
4. <https://www.microsoft.com/ro-ro/security/business/microsoft-defender> [↑](#footnote-ref-4)
5. <https://www.proofpoint.com/us/solutions/protect-against-phishing> [↑](#footnote-ref-5)
6. <https://www.kaggle.com/datasets/naserabdullahalam/phishing-email-dataset/data> [↑](#footnote-ref-6)
7. <https://www.webroot.com/us/en/resources/tips-articles/spam-vs-phishing?srsltid=AfmBOorzQvHxZ612MWLeUYw2GeUe00ApoiHpSam5uJ7Y-OwtE4SUbvfq> [↑](#footnote-ref-7)
8. <https://monkey.org/~jose/phishing/> [↑](#footnote-ref-8)
9. <https://www.kaggle.com/datasets/rtatman/fraudulent-email-corpus> [↑](#footnote-ref-9)
10. <https://www.kaggle.com/datasets/beatoa/spamassassin-public-corpus> [↑](#footnote-ref-10)
11. <https://www.cs.cmu.edu/~enron/> [↑](#footnote-ref-11)