**Laborator 4 – Proiect de cercetare**

**Sistem de detectare a phishing**

**Ce date se folosesc?**

* Se utilizează un set de date diversificat, cuprinzând e-mailuri reale și simulări de atacuri de phishing. Acestea sunt colectate din baze de date publice. Se iau în considerare diverse formate de e-mail, atașamente și link-uri.

**Ce experimente se fac?**

* Experimentele includ antrenarea modelului pe setul de date menționat, evaluarea performanței acestuia pe un set de testare, și compararea rezultatelor cu alte abordări existente. Se efectuează experimente pentru analiza robusteței modelului în fața diverselor tactici de phishing, cum ar fi variații în textul e-mailului și tehnicile de camuflare a link-urilor.

**Cum validăm rezultatele?**

* Validarea rezultatelor se realizează prin metrici precum precizia si revocare. Se folosesc tehnici precum cross-validation pentru a asigura robustețea și generalizarea modelului.

**Cu ce trebuie să comparăm rezultatele pentru a arăta că abordarea originală aduce beneficii față de abordările existente în literatură?**

* Rezultatele sunt comparate cu abordările existente în literatură, inclusiv modele de învățare automată și tehnici de analiză a semnăturilor phishing. Se pune accent pe îmbunătățirile aduse în ceea ce privește precizia, revocarea și capacitatea de adaptare la noi tactici de phishing, demonstrând astfel avantajele abordării propuse.

**Creare model matematic/riguros care descrie cu exactitate experimentele propuse.**

* Modelul propus este o rețea neurală recurentă (RNN) cu straturi multiple, antrenată cu un algoritm de optimizare specific. Fiecare strat al rețelei are funcții de activare și parametri specifici, iar antrenarea este efectuată prin minimizarea unei funcții de cost definite. Detaliile matematice exacte, precum și justificările pentru arhitectura aleasă, sunt incluse în documentația modelului.