
UNIVERSITATEA BABEŞ-BOLYAI UNIVERSITY, CLUJ-NAPOCA, ROMÂNIA
FACULTATEA DE MATEMATICĂ ȘI INFORMATICĂ

# Detecția și analiza linkurilor web malițioase

–Rezumatul tezei de doctorat–

**Autor**
Claudia-Ioana Coste
Conducător de doctorat: Prof. Univ. Dr. Anca-Mirela Andreica

2026

# Cuprins

# Capitolul 1

# Introducere

Ultimele două decenii au fost marcate de dezvoltarea exponențială a Internetului. În zilele noastre, viața fără Internet nu mai poate fi posibilă. Conform Statista [26], există 5,52 miliarde de utilizatori de Internet la nivel mondial, ceea ce reprezintă aproximativ 67,31% din populația globală (8,2 miliarde [31]). Utilizatorii obișnuiți navighează pe Internet pentru a-și realiza sarcinile și îndatoririle zilnice, de exemplu: cumpărături, socializare, administrație publică, educație, divertisment și așa mai departe. Cu toate acestea, oamenii nu pot ține pasul cu ritmul în care se dezvoltă tehnologia web. Astfel, mulți utilizatori au puține cunoștințe digitale și constituie ținte vulnerabile pentru diverse atacuri, care devin din ce în ce mai sofisticate.

Majoritatea amenințărilor online se bazează pe phishing, având ca scop păcălirea consumatorilor și determinarea acestora să dea click pe linkuri. Acestea pot să descarce programe malware fără permisiunea lor, să le fure informațiile personale și să propage conținut fals pentru a le influența comportamentul. Legăturile web malițioase sunt acele site-uri web stocate pe World Wide Web (WWW), care desfășoară orice fel de comportament rău-intenționat. De exemplu, linkurile maligne pot găzdui site-uri clonate utilizate pentru a induce în eroare utilizatorii să își introducă datele de autentificare. Mai mult decât atât, unele Uniform Resource Locator (URL)-uri rulează cod malițios, descărcând programe fără consimțământul utilizatorului sau pot urmări acțiunile consumatorilor online, pot rula diverse sarcini în browser, care consumă resursele computaționale locale etc. În plus, site-urile web malițioase pot conține articole jurnalistice de calitate scăzută (adică clickbait) sau știri complet false, contribuind la fenomenul de dezinformare.

De asemenea, aceste site-uri create cu rea-intenție sunt propagate în mediul online prin clickbait, mesaje scurte, exagerate și dramatice sau create în mod special pentru a induce în eroare consumatorii și a-i determina să acceseze pagina web și să o distribuie. De obicei, clickbait-ul este folosit în principal cu articole jurnalistice de calitate indoielnică pentru a obține cât mai multe vizualizări, clickuri și beneficii

financiare mai mari din publicitate. Pe lângă câștigurile financiare, clickbait-ul poate fi utilizat pentru a răspândi știri false, pentru a propaga informații false și a influența comportamentul utilizatorilor. Știrile false sunt știri care conțin informații false și totodată sunt create cu intenția vicioasă de a amăgi cititorii [27].

## 1.1 Motivație

Conform Insider Intelligence [25], se estimează că utilizatorii de Internet petrec aproximativ patru ore pe zi pe dispozitivele mobile, utilizând diverse aplicații de social media și Internetul în general. Din această cauză, aceștia pot fi ținte vulnerabile pentru diverse atacuri malware. Din păcate, conform statisticilor europene [22], alfabetizarea online este la un nivel destul de scăzut în Uniunea Europeană, cu o medie de 60%. Acest procent reprezintă indivizii cu abilități digitale de bază și mediocre. Clasamentul discutat în [22] plasează România și Bulgaria pe ultimele poziții, cu aproape 25% din indivizi având încredere în abilitățile lor digitale. Mediul online evoluează într-un ritm mult mai alert decât sunt oamenii capabili să își dezvolte noi abilități și să se adapteze la mediul online. Astfel, atacurile web devin din ce în ce mai sofisticate, iar sarcina de a le identifica a devenit din ce în ce mai anevoioasă. Conform Forbes [25], în 2023, erau disponibile 1,11 miliarde de URL-uri pe Internet, dintre care 202 milioane erau active, furnizând diverse resurse și servicii. În plus, compania AVG raportează că aproape un milion de site-uri web unice de phishing au fost create doar în primul trimestru al anului 2024 [19]. Principalul vector de propagare pentru malware-ul web este reprezentat de linkurile malițioase adesea integrate în mesaje înșelătoare distribuite prin rețele de socializare și alte platforme [19]. Alți vectori de propagare pentru amenințările web sunt paginile web deja infectate. De exemplu, în cazul website-urilor create cu WordPress, există estimări care indică faptul că aproximativ 700.000 de pagini WordPress conțineau cel puțin un fișier malițios [19]. WordPress este unul dintre cele mai populare sisteme de gestionare a conținutului, iar aproape 40% din toate site-urile web sunt create folosind acest sistem [19].

Linkurile web malițioase pot fi asociate cu o varietate largă de amenințări online, cum ar fi redirecționarea malițioasă, execuția de cod pentru a utiliza resursele browserului, *drive-by-downloads* (programe malware care se descarcă automat), site-uri web clonate și site-uri phishing etc. În plus, acestea pot distribui știri false și articole de o calitate îndoielnică, dar cu titluri foarte atractive.

Clickbait este definit de dicționarul Oxford ca material disponibil pe Internet, care a fost creat cu scopul de a atrage atenția și de a convinge consumatorii să acceseze linkul web care indică către un site web anume [23]. Clickbait poate denumi și articole jurnaliste sub standardul minim de calitate,

care promit ceva din titlu, dar îi amăgesc pe cititori. Scopul folosirii tehnicilor de tip clickbait este să genereze clickuri și vizualizări astfel încât să se obțină un câștig financiar cât mai mare din publicitate. Există multiple tehnici folosite pentru a atrage atenția, iar multe dintre ele sunt utilizate în mass-media jurnalistică, la televizor sau în presa scrisă. Conținutul acestor articole nu corespunde cu așteptările stabilite prin titlu și poate prezenta informații false sau pseudostiință.

Clickbait poate fi folosit în continuare pentru a contribui la distribuirea de știri false. Conform [27], știrile false sunt știri care prezintă informații factual greșite care pot fi dovedite ca atare. În plus, acestea sunt fabricate cu intenția malițioasă de a păcăli cititorii. Știrile false s-au dovedit că au influențat comportamentul votanților la alegerile SUA din 2016 și la referendumul Regatului Unit privind Brexit-ul. De exemplu, în Marea Britanie în timpul organizării referendumului privind ieșirea din Uniunea Europeană (UE), campania contra-UE avea mai mulți urmăritori și era distribuită mai des decât cea pro-UE, conform [29]. De asemenea, din cauza lipsei unor reglementări corespunzătoare, știrile false influențează comportamentul electoral al consumatorilor, așa cum s-a întâmplat în Statele Unite în 2016. În [17], este menționat că articolul cel mai distribuit în 2016 a fost o postare de pe un blog intitulată „De ce votez cu Donald Trump", scrisă de o cunoscută femeie conservatoare. Strategia pare să se fi repetat și în timpul alegerilor din 2024 din Statele Unite și în alte țări (de exemplu, România). În România, rezultatele rundei întâi de alegeri au fost anulate din cauza influenței rusești dovedite prin răspândirea de informații înșelătoare pe TikTok [16]. Adesea, știrile false sunt popularizate prin rețele de socializare, folosind troli, boți, site-uri malițioase, clickbait și altele. În ansamblu, știrile care prezintă informații false pot fi folosite pentru a influența consumul, sănătatea, comportamentul sau stilul de viață al cititorilor.

## 1.2 Obiective

Teza de față ia în considerare domeniul de detecție al linkurilor web malițioase din mai multe perspective. Aceasta elaborează și urmărește contribuții aduse la mai multe etape ale pipeline-ului pentru clasificarea linkurilor web malițioase. În primul rând, procesul de extragere a caracteristicilor este investigat și există o perspectivă experimentală asupra analizei importanței atributelor folosite în modelare. Mai mult decât atât, se urmărește extragerea de informații din linkurile web și transformarea URL-urilor în imagini. În al doilea rând, o gama variată de algoritmi de învățare automată sunt aplicați, arhitecturi noi sunt propuse împreună cu comparații relevante. În al treilea rând, linkurile web malițioase sunt analizate luând în considerare un spectru mai larg de atacuri web înșelătoare, precum clickbait-ul și știrile false. De asemenea, este important de menționat studiul realizat pentru

articolele de literatură publicate în acest domeniu al detecției site-urilor web maligne. Lucrările au fost analizate și clasificate conform datelor culese, caracteristicilor utilizate și algoritmilor de clasificare. În plus, au fost efectuate comparații relevante între mai multe abordări.

Obiectivele pot fi formulate considerând mai multe perspective, legăturilor web malițioase, click-baitul și știrile false. Prin urmare, obiectivele propuse pe direcția linkurile web malițioase sunt următoarele:

O1) Revizuirea literaturii de specialitate pentru detecția de linkuri web malițioase [7];

Pentru acest obiectiv, analiza trebuie să urmărească pipeline-ul de detecție, incluzând setul de date, colectarea datelor, procesul de adnotare, atributele folosite la modelare și algoritmul de clasificare. Atributele extrase și metodele de clasificare sunt categorisite și împărțite în grupuri. De asemenea, se include și o scurtă descriere a fiecărei abordări.

O2) Experimentarea cu numeroși algoritmi de învățare automată [10], [4], [3], [11], [13] și [12];

Testele realizate ar trebui să considere o mare diversitate de metode de inteligență artificială precum: K-Nearest Neighbor (KNN), Random Forest (Pădure aleatoare de arbori) (RF), Decision Tree (Arbore de decizie) (DT), Logistic Regression (Regresie Logistică) (LR), Naive Bayes (NB), Adaptive Boosting (ADA), Gradient Boosted Tree (XGB), Support Vector Machine (Mașină cu suport vectorial) (SVM) și alte tehnici de învățare automată, diferite modele ansamblu, de exemplu ansambluri cu vot majoritar [4], ansambluri cu vot majoritar ponderat [13], [12] și ansamblu bazat pe Particle Swarm Optimization (PSO) [3], arhitecturi pentru învățarea profundă precum: Convolutional Neural Network (Rețea neuronală convoluțională) (CNN), Long Short Term Memory (Mecanism de memorie de scurtă și de lungă durată) (LSTM), ResNet101, ResNet50V2, VGG16 [11] și diverse modele mari de limbaj [12].

Mai mult decât atât, este necesară și propunerea unei metodologii solide pentru derularea de experimente [10].

O3) Folosirea mai multor seturi de date [3], [11];

Experimentele ar trebui să integreze mai multe seturi de date, de diverse tipuri.

O4) Experimentarea cu diferite categorii de atribute;

Prin folosirea mai multor categorii de caracteristici, mai multe metode de detecție a linkurilor malițioase pot fi propuse. De exemplu, atribute care sunt extrase automat prin învățare automată [11] sau formule [3], sau extrase manual [13] și [12]. Atributele extrase pot trece prin transformări și preprocesări ulterioare pentru a îmbunătății modelul de clasificare.

Obiectivele proiectate pentru detecția de clickbait și știri false sunt următoarele:

O5) Investigarea viabilității colectării automate de legături web non-clickbait folosind logurile serverului Squid (un server proxy web) [1];

O6) Experimentarea cu multiple tehnici de învățare automată, de exemplu: RF în [2] și [9], LR, NB, DT, SVM atât pentru detecția clickbait cât și pentru clasificarea știrilor false [6], [5], [8];

O7) Extragerea manuală de atribute care sunt independente de limbă [6], [5], [9], [8];

O8) Ierarhizarea celor mai importante caracteristici folosite la clasificarea clickbait-ului și a știrilor false [6], [5], [9], [8];

O9) Investigarea asocierii și a relației existente între clickbait, știri false și linkurile malițioase ca vector de distribuție.

## 1.3 Contribuții originale

Contribuțiile originale ale tezei pot fi împărțite în patru domenii, considerând problema abordată. Categoriile sunt: detecția de linkuri web malițioase, detecția titlurilor înșelătoare din mediul online, detecția de știri false și în final, relația dintre linkuri, clickbait și articolele care conțin informații false.

În primul rând, pentru domeniul linkurilor web malițioase, putem contura următoarele contribuții:

- Recenzia celor mai importante soluții propuse pentru clasificarea linkurilor [7];
  - Clasificarea și descrierea metodelor propuse în literatura de specialitate;
  - Categorisirea și prezentarea caracteristicilor;
  - Analiza comparativă și discuțiile referitoare la reproductibilitatea studiului, avantajele și dezavantajele metodelor propuse etc.
- Proiectarea unei noi metodologii pentru calibrare și pentru comparații [10] utilizând setul de date propus în [30];
  - Rafinarea unei metodologii pentru detecția linkuri web maligne care poate fi folosită la experimentarea cu algoritmi de învățare automată;
  - Folosirea metodologiei propuse anterior pentru antrenarea și compararea mai multor modele de învățare automată precum RF, DT și KNN. Cel mai bun algoritm a fost RF cu 87,6% precizie;

  - Caracteristicile extrase de tip *host-based* s-au dovedit mai relevante decât cele lexicale;

- Dezvoltarea unui ansamblu bazat pe votul majorității pentru clasificarea linkurilor [4] folosind setul de date pus la dispoziție de [30];
  - Compararea și calibrarea a zece metode de învățare automată (LR, NB, Ada Boost, XGB, Linear Discriminant Analysis (LDA), Multi-layer Perceptron (Perceptron cu straturi multiple) (MLP) și SVM);
  - Cel mai bun model individual a fost XGB cu 95,07% precizie. Ansamblurile cu vot majoritar au fost construite în toate combinațiile posibile (120 de variațiuni) și cel mai bun (ansamblul Adaptive Boosting - Gradient Boosted Tree - Support Vector Machine with Radial Basis Function (ADA-XGB-SVM-rbf)) a reusit să atingă o precizie de 96,31%. Acesta a surclasat modelele individuale și alte soluții folosite pentru comparație;
- Propunerea unui nou ansamblu bazat pe algoritmul *Particle Swarm Optimization* folosit pentru detecția linkurilor suspicioase [3];
  - Compararea a 12 modele singulare de învățare automată, din care amintim: LR, SVM, ADA, RF, DT, KNN, Perceptron, Nearest Centroid (NC), Passive Aggressive Classifier (Clasificator pasiv-agresiv) (PAC), Stochastic Gradient Descent (SGD), *KMeans* și diferite variante pentru NB;
  - Ansamblul bazat pe PSO este construit;
  - Testări pe două seturi de date: D1 [21] și D2 [28];
  - Pentru D1 [21], modelul cu PSO a atins o acuratețe de 97,05% și pentru D2 [28], s-a ajuns la 91,12% acuratețe;
  - Rezultatele au depășit performanța atinsă de alte metode din literatură;
- O abordare pentru folosirea modelelor mari de limbaj pentru clasificarea linkurilor în [13] și [12];
  - Îmbogățirea modelului de clasificare cu caracteristici extrase manual pe datasetul D2 [28] cu o etichetă generată cu modelul de la OpenAI (GPT-4 și GPT-2 calibrat);
  - Experimente cu multiple modele mari de limbaj (modelele Generative Pre-trained Transformer (Transformator generativ pre-antrenat) (GPT): GPT-3.5-turbo, GPT-4, o1-mini, GPT-4o-mini și GPT-2 calibrat);
  - Compararea a 13 tehnici de învățare automată și a ansamblurilor cu vot majoritar ponderat pentru a observa dacă adnotarea făcută cu GPT conduce la o performanță mai bună;

  - Experimentele au dovedit o ușoară îmbunătățire pentru majoritatea algoritmilor;

- Elaborarea unei abordări inovative prin transformarea URL-urilor în imagini și aplicarea de metode de învățare profundă pentru detecția de linkuri suspicioase [11];
  - Experimentarea cu două seturi de date D1 [21] și D2 [28];
  - Generarea a două tipuri de imagini: color și alb-negru;
  - Multiple configurații de arhitecturi pentru modelele de învățare profundă aplicate (CNN, LSTM, VGG16, ResNet50V2, ResNet101) și calibrare;
  - Acuratețe de 96,82% pentru D2 [28] surclasând metodele echivalente;
  - Nu sunt diferențe semnificative pentru clasificarea folosind imagini color și cele alb-negru;
  - LSTM-ul poate fi influențat de modalitatea prin care informația a fost codificată în imagine, dacă s-a făcut iterând rândurile prima dată sau coloanele.

În al doilea rând, pentru detecția de clickbait, mai multe lucruri au fost realizate:

- Colectarea de exemple non-clickbait bazată pe euristici aplicate pe șablonul de activități online ale utilizatorilor creat din loguri Squid [1];
- Propunerea unei strategii independente de limbă pentru detecția clickbait folosind doar caracteristici extrase manual [2, 6, 5];
- Mai multe experimente rulate pe un dataset de dimensiune mare în limba engleză antrenând mai mulți algoritmi de învățare automată, precum RF [6], LR, NB, SVM, luând în considerare diferite categorii de proprietăți [5];
- Obținerea unei performanțe cu 73,93% acuratețe pentru LR și SVM, considerând atribute lexicale și o performanță de 71,81% pentru RF, luând în calcul toate categoriile de caracteristici [5];
- Dezvoltarea unei analize privind importanța caracteristicilor, care a reușit să confirme mai multe rezultate deja obținute în literatură pentru detecția clickbait [2, 6, 5].

Luând în seamă problema detecției știrilor false, câteva contribuții au fost obținute și în [8].

- Propunerea unui model de clasificare bazat pe caracteristici;
- Teste pe două seturi de date în limbi diferite (engleză și germană);
- Compararea mai multor modele de învățare automată, cum ar fi: RF, LR, NB și SVM cu kernel liniar și bazat pe Radial Basis Function (RBF);

- Obținerea unei acurateți de 75,38% pentru modelul RF;
- Analiza celor mai relevante caracteristici.

Într-un final, relația dintre legăturile web malțioase, clickbait și știri false a fost explorată și rezultatele analizei ar putea fi rezumate la următoarele:

- Clickbait-ul este o tehnică jurnalistică care contribuie la răspândirea de știri false și la alte tipuri de articole cu un conținut de calitate inferioară;
- Clickbait-ul și știrile false prezentate într-un context european și global [2];
- Linkurile web pot găzdui știri false și articole de o calitate jurnalistică slabă, prin urmare aceste linkuri pot fi considerate ca fiind malițioase.

## 1.4 Listă de publicații

Standardele de evaluare sunt cele valide la data de 1 Octombrie 2018. Lista de conferințe precum și categorisirea lor este bazată pe clasificarea internațională CORE. Lista de jurnale este cea folosită de UEFISCDI pentru premierea articolelor publicate în jurnalele științifice internaționale.

1. Kiraly, A., **Coste, C-I.**, Șotropa, D-F. and Bufnea, D.-V. (2026). Towards Efficient Phishing Email Detection with LLM-Based Models. In Proceedings of the Intelligent Decision Technologies Conference (KES-IDT-26). KES Smart Innovation Systems and Technologies. Accepted for publication. In press. - Publicație indexată în Scopus și Web of Science, rang conferință C, puncte: 2/(4-2) = 1;
2. **Coste, C-I.**, Chira, C. and Bufnea, D.-V. (2026). Malicious Web Links Detection: A Survey. SUBMITTED. - în proces de review, puncte: 0;
3. **Coste, C-I.** and Kaisser, T. (2026). Malicious Web Links Detection with Large Language Models. Lecture Notes in Business Information Processing. Accepted for publication. In press. - jurnal indexat în Scopus, dar nu a fost găsit în listele UEFISCDI, puncte: 1;
4. Moldovan, D.-A. and **Coste, C-I.** (2025). Music Emotion Recognition with Deep Learning Techniques. In: Chira, C., Matei, O., Pop, F., Pop-Sitar, P. (eds) Innovative Perspectives on Computational Intelligence and Data Science. InnoComp 2025. Communications in Computer and Information Science, vol 2793. Springer, Cham. 10.1007/978-3-032-12478-4_11 - jurnal indexat în Scopus, dar nu a fost găsit în listele UEFISCDI, puncte: 1;

5. **Coste, C-I.** (2024). Malicious Web Links Detection Based on Image Processing and Deep Learning Models. In Proceedings of the 2024 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), ISBN 979-8-3315-0494-6/24, pages 445-449. - rang conferință B (lucrare scurtă), puncte: 2/3*4 = 2.66;

6. Kaisser, T. and **Coste, C-I.** (2024). Using Chat GPT for Malicious Web Links Detection. In Proceedings of the 20th International Conference on Web Information Systems and Technologies, ISBN 978-989-758-718-4, ISSN 2184-3252, pages 425-432. - Premiată cu Best Student Paper - rang conferință C (lucrare scurtă), puncte 2/3*2 = 1.33;

7. **Coste, C-I.** (2024). Using Ensemble Models for Malicious Web Links Detection. In Proceedings of the 16th International Conference on Agents and Artificial Intelligence - Volume 3, ISBN 978-989-758-680-4, ISSN 2184-433X, pages 657-664. - rang conferință B (lucrare scurtă), puncte: 2/3*4 = 2.66;

8. **Coste, C-I.**, Andreica, A-M., and Chira, C. (2023). Malicious Web Links Detection Using Ensemble Models. In Proceedings of the 19th International Conference on Web Information Systems and Technologies, ISBN 978-989-758-672-9, ISSN 2184-3252, pages 268-275. - rang conferință C (lucrare scurtă și prezentare sub formă de poster), puncte: 2/3*2 = 1.33;

9. **Coste, C-I.** (2023). Malicious Web Links Detection - A Comparative Analysis of Machine Learning Algorithms. Studia Universitatis Babeș-Bolyai Informatica, ISSN 2065-9601. [S.l.], v. 68, n. 1, pages 21-36. - jurnal BDI, puncte: 1;

10. **Coste, C-I.**, and Bufnea, D.. (2021). Advances in Clickbait and Fake News Detection Using New Language-independent Strategies. Journal of Communications Software and Systems 17.3, pages 270-280. - jurnal indexat în Scopus, Se regăsește în Q4 (cuartila 4) conform listei puse la dispoziție de UEFISCDI, puncte: 2;

11. **Coste, C-I.**, Bufnea, D., and V. Niculescu. (2020). A new language independent strategy for clickbait detection. In Proceedings of the International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Hvar, Croatia, pages 1-6, 10.23919/SoftCOM50211.2020.9238342; - rang conferință D, puncte: 1;

12. **Coste, C-I.** (2019). Online bad habits: fake news and clickbait. In Proceedings of the Student Interdisciplinary Conference The European Union and Global Order, Cluj University Press, pp. 179-193, April 5th, 2019, Cluj-Napoca, Romania; - jurnal, puncte: 0;

13. **Coste, C-I.** (2018). Controlling the click bait. In Proceedings of the International Student Conference StudMath-IT, ISSN 2602-1579, pages 11-18, Arad, 15th - 16th November 2018 - Conferință studențească, puncte: 0.

Număr total de puncte: 13 puncte

## 1.5 Rezumat

Rezumatul tezei de față este structurat în șase capitole.

Primul capitol prezintă o introducere în nișa de cercetare abordată împreună cu motivația pentru alegerea făcută. În plus, sunt incluse obiectivele tezei, contribuțiile originale cele mai importante, rezumatul și lista de publicații.

Al doilea capitol parcurge stadiul actual al cunoașterii în domeniu raportat la cele mai importante soluții pentru detecția linkurilor web malițioase. Capitolul discută ideile principale din articolele citite, împărțindu-le în două grupuri, soluții dinamice și statice.

Al treilea capitol, intitulat „Comparație sistematică și analiza tehnicilor de detecție a linkurilor web malițioase", discută și oferă o analiză critică a literaturii de specialitate, menționând câteva observații generale despre fluxul utilizat pentru detecția linkurilor suspicioase.

Capitolul următor include contribuțiile principale, prezentând cele mai relevante soluții propuse în teză. Sunt relatate informații despre seturile de date, preprocesare, despre diferitele metode aplicate pentru extragerea caracteristicilor, împreună cu larga varietate de algoritmi de clasificare.

Penultimul capitol prezintă investigația privind clickbait-ul și știrile false, incluzând colectarea datelor folosind euristici și loguri Squid, algoritmi de clasificare și analiza importanței caracteristicilor.

În final, ultimul capitol evidențiază concluziile, dacă obiectivele propuse au fost atinse și contribuțiile principale.

# Capitolul 2

# Abordări precedente în detecția de linkuri web malițioase

Capitolul de față își propune să discute mai multe tipuri de soluții privind detecția de linkuri web malițioase. Sunt trecute în revistă metode precum *honeypots*, abordări bazate pe *sandbox*-uri, antiviruși, soluții dezvoltate cu liste cu site-uri blocate, rețele complexe și învățare automată.

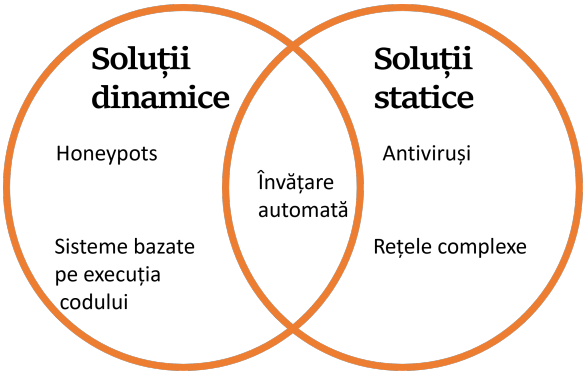


Figura 2.1: Clasificarea articolelor studiate conform [7].

Prin studierea majorității articolelor din literatură care consideră detecția de linkuri web malițioase, s-a observat că toate soluțiile ar putea fi împărțite în două categorii: soluții dinamice și statice precum în [18]. În categoria soluțiilor dinamice, se pot include abordări care permit accesarea linkului, rularea codului (Hypertext Markup Language (HTML), Cascading Style Sheets (CSS), sau cod JavaScript) și analiza comportamentului. Categoria soluțiilor dinamice include *honeypots*, soluții cu *sandbox*-uri și modele de învățare automată care aplică caracteristici extrase pe baza comportamentului. Abordările care folosesc *honeypot*-uri își propun să atragă și să inducă în eroare linkuri malițioase, astfel încât acestea să fie analizate într-un mediu sigur unde pot fi executate. Studiile folosind doar medii de

execuție (*sandbox*-uri) pentru detecție au scopul de a lăsa pagina web să fie accesată și executată într-un mediu securizat și comportamentul acesteia să poată fi observat. Pe baza acestui comportament, legătura web va fi marcată ca fiind suspicioasă sau nu. Soluțiile statice previn accesarea linkului și ele adnotează linkurile luând în considerare doar proprietăți statice. Din categoria soluțiilor statice, putem aminti metode de învățare automată, programe antivirus și analiza rețelelor complexe. Asemănarea dintre cele abordări, static și dinamic, în cazul soluțiilor bazate pe învățare automată este faptul că ambele extrag atribute care pot fi folosite mai târziu ca date de intrare la modelele inteligente [7].

O reprezentare vizuală sugestivă pentru clasificarea rezolvărilor la detecția de linkuri web malițioase poate fi vizualizată în Figura 2.1. Ca observație, sunt multe articole de specialitate care propun idei ce pot fi incluse în una sau mai multe categorii din cele discutate mai sus. De exemplu, sunt programe antivirus care combină abordarea bazată pe semnături cu euristici relevante și chiar algoritmi inteligenți. Mai mult decât atât, sunt metode care folosesc rețele complexe combinate cu alte metode inteligente [7].

# Capitolul 3

# Comparație sistematică și analiza tehnicilor de detecție a linkurilor web malițioase

Capitolul curent propune o analiză critică și o clasificare a soluțiilor existente până la momentul de față în domeniul detecției linkurilor web malițioase. Toate abordările vor fi studiate conform etapelor descrise în [7]. Aceste etape includ observarea setului de date, modalitatea de extragere și selecția caracteristicilor, algoritmii de clasificare utilizați, reproductibilitatea lucrărilor și comparații relevante.

Luând în considerare cercetarea din [7], majoritatea abordărilor folosesc seturi de date deja disponibile, precum și agregarea unora dintre ele cu scopul obținerii unui set de date mai mare. Sistemele de detecție care utilizează un singur set de date au folosit majoritar dataseturi disponibile pe [20]. În plus, aproximativ jumătate din articolele de cercetare investigate au folosit un set de date privat, care nu a fost făcut disponibil public. Cu toate acestea, în majoritatea cazurilor, articolele de specialitate își specifică sursele [7].

Soluțiile care aleg să-și colecteze propriile linkuri și proprietățile asociate cu un URL folosesc un web crawler pentru a prelua informațiile de pe Internet [7]. Multe înregistrărilor benigne sunt preluate de pe site-uri care fac un clasament al linkurilor web accesate, precum Alexa's Top [15]. În plus, alte URL-uri benigne sunt colectate prin motoare de căutare, așa cum se specifică în [14]. Linkurile malițioase sunt adesea preluate de la listele publice care conțin site-uri blocate.

Procesul de adnotare este realizat folosind *sandbox*-uri și interfețe furnizate de programele antivirus sau alte servicii de ierarhizare pentru website-uri.

Luând în considerare atributele utilizate în modelarea clasificării și construirea modelului inteligent,

de obicei, abordările iau în considerare doar URL-ul. Astfel, multe studii se concentrează mai mult pe caracteristici independente de conținut. Totuși, există și alte metode care consideră relevant conținutul unei pagini web, în special codul JavaScript (JS) și HTML.

Procesul de extragere a caracteristicilor reprezintă etapa cea mai importantă și cea care oferă cunoștințe în procesul de clasificare. Linkul web împreună cu conținutul acestuia, trebuie transformat în numere, astfel încât algoritmului de detecție să primească aceste informații ca date de intrare. Procesul de extragere a caracteristicilor poate fi realizat fie automat, folosind instrumente de Machine Learning (Învățare automată) (ML) (de exemplu, autoencodere, rețele neuronale convoluționale, etc.), fie prin inginerie manuală. Ingineria manuală se bazează pe euristici și informații din domeniu puse la dispoziție de experți, astfel, anumite caracteristici mai relevante sunt selectate și extrase. În plus, informațiile importante pot fi extrase și prin alte metode: rețele complexe, grafuri, formule matematice, codificare binară etc.

În ceea ce privește reproductibilitatea lucrărilor analizate, doar aproximativ cinci din toate articolele analizate (mai mult de 50) au oferit suficiente detalii de implementare și au pus la dispoziție codul pe care l-au folosit pentru experimente. Setul de date a fost furnizat pentru aproximativ 37% din lucrările cercetate conform [7].

În plus, au fost făcute mai multe comparații relevante între diferite tipuri de soluții, considerând efectele atacurilor, explicabilitatea algoritmului de detecție, atacurile de tip *0-day*, evoluția în timp a legăturilor web, tehnici folosite de dezvoltatori pentru a își proteja codul, resursele computaționale și capacitatea de stocare necesară.

# Capitolul 4

# Abordări propuse pentru detecția de linkuri web malițioase

Acest capitol își propune să prezinte succint abordările propuse privind detecția linkurilor web malițioase.

În primul rând, sunt descrise seturile de date utilizate. Pentru toate soluțiile prezentate în această teză, au fost folosite trei seturi de date. Setul inițial de date [30] are caracteristici *host-based* și lexicale deja extrase. Al doilea și al treilea set de date sunt mai mari, dar conțin doar URL-uri și eticheta. Primul set de date (D1) poate fi găsit la [21] și al doilea (D2) la [28].

În continuare, metodele de preprocesare sunt detaliate. Pentru setul de date Urcuqui et al. [30], preprocesarea a inclus patru etape conform lucrării [10]: curățare; adaptarea datelor; codarea caracteristici categorice și normalizarea datelor. De asemenea, preprocesarea în [11] a constat în transformarea URL-urilor în imagini luând în considerare codul American Standard Code for Information Interchange (Codul standard american pentru schimb de informații) (ASCII) care a fost tradus într-o nuanță alb-negru sau color. Pentru experimentele prezentate în lucrarea [3], setul de date D2 [28] a fost preprocessat cu formula matematică Term Frequency and Inverse Document Frequency (TF-IDF), implementată în biblioteca Python Sklearn [24].

În etapa următoare, metodologiile empirice sunt prezentate. Algoritmii propuși sunt descriși în detaliu. Multiple metode de învățare automată sunt introduse, cum ar fi *K-Nearest Neighbor*, Arbore de decizie, Pădure aleatoare de arbori, *Adaptive Boosting*, Regresie Logistică, diferite variante pentru Naive Bayes, Mașina cu suport vectorial, *Linear Discriminant Analysis*, *Quadratic Discriminant Analysis*, *Gradient Boosted Tree*, *Multi-layer Perceptron* etc. În plus, o strategie pentru calibrarea metodelor și exprimentare a fost dezvoltată și aplicată pentru trei algoritmi de învățare automată după cum este detaliat în [10].

De asemenea, trei tipuri de modele ansamblu sunt propuse. Rezultatul clasificării se bazează pe detecția algoritmilor individuali. Un model ansamblu cu vot majoritar este propus în [4], unde clasa finală este aleasă pe baza majorității adnotărilor tuturor clasificatorilor. Ansamblurile cu coeficienți generați cu *Particle Swarm Optimization* sunt discutate în [3], unde fiecare coeficient e legat de un clasificator care formează modelul hibrid. Un alt tip de ansamblu cu vot majoritar ponderat este dezvoltat folosind o caracteristică extrasă de un model mare de limbaj în [13] și în [12]. Fiecărei adnotări făcute de fiecare clasificator în parte îi este asociat un coeficient, o rată de încredere. Clasa finală este calculată ca media ponderată a tuturor etichetelor care au fost generate de clasificatori. În [13], au fost utilizate două modele de la OpenAI, cum ar fi GPT 3.5-turbo și GPT 4. În [12], două modele GPT mini au fost adăugate pentru teste: GPT-4o-mini (GPT-4o-mini-2024-07-18) și o1-mini (o1-mini-2024-09-12).

De asemenea, soluții care includ învățarea profundă sunt luate în considerare pentru clasificarea imaginilor unde informația conținută de URL este codificată [11]. Mai multe arhitecturi au fost propuse, cum ar fi o Rețea neuronală convoluțională cu trei straturi convoluționale și alte arhitecturi cunoscute precum ResNet50V2, ResNet101 și VGG16. În final, modelele au fost îmbogățite cu un strat LSTM cu intenția de a îmbunătăți performanța și de a capta mai bine informația încifrată în imagine.

În final, în restul capitolului, experimentele și rezultatele sunt prezentate. Acestea sunt comparate și discutate critic prin raportare cu abordările similare din literatură.

# Capitolul 5

# Linkuri web malițioase în asociere cu clickbait (titluri înșelătoare) și știri false

În cele ce urmează, este prezentată perspectiva curentă a tezei asupra relației dintre legăturile web malițioase, clickbait și știrile false. Mai întâi, domeniul de detecție a titlurilor înșelătoare din mediul online este descris, împreună cu cercetările și experimentele propuse de această lucrare. Propunerea noastră privind procesul de colectare și adnotare a datelor este prezentată în [1] și [9], unde a fost generată o euristică privind etichetarea automată a exemplelor de site-uri non-clickbait folosind șabloane de navigare ale utilizatorilor create pe baza jurnalelor de sistem înregistrate de serverul web proxy Squid. În plus, ideile noastre în detecția clickbait-ului, care iau în considerare caracteristici independente de limbă și au fost extrase manual, sunt detaliate în [2], [9], [6], [5] și [8]. Metodologia folosită a inclus mai întâi detectarea limbii, iar apoi extragerea caracteristicilor din mesajul clickbait sau din titlul știrii. De asemenea, domeniul detecției știrilor false este prezentat cu experimentele propuse, unde abordările de învățare automată sunt testate și se realizează o analiză a importanței atributelor folosite în procesul de modelare [8]. Mai mulți algoritmi inteligenți au fost folosiți pentru detecția atât a știrilor false cât și a clickbait-ului, printre care LR, RF, NB, DT și SVM.

În ultima secțiune a capitolului curent, relația dintre clickbait și știrile false este explicată luând în considerare contextul global și european conform [2], iar ulterior este discutată asocierea lor cu linkurile web malițioase.

# Capitolul 6

# Concluzii și direcții viitoare de cercetare

Acest capitol va descrie principalele concluzii din această teză și va indica câteva direcții de cercetare viitoare.

## 6.1 Concluzii

Detecția legăturilor web malițioase este un domeniu complex și dificil, iar teza dorește să avanseze cercetarea în acest domeniu. Investigația a început prin recenzia celor mai relevante lucrări scrise în domeniu. Toate soluțiile și caracteristicile utilizate în sistemele de detecție au fost clasificate și descrise. Mai multe comparații au fost făcute între diverse tipuri de soluții și între soluții, atunci când acest lucru era potrivit. De asemenea, au fost discutate alte observații, precum reproductibilitatea lucrării.

În ceea ce privește obiectivele propuse pentru detecția linkurilor web malițioase, acestea au fost atinse în cea mai mare parte. O revizuire sistematică a literaturii a fost făcută cu multiple analize asupra seturilor de date utilizate, caracteristicilor și unei game largi de soluții de detecție (obiectivul O1). Mai multe categorii de algoritmi de inteligență artificială au fost testați pentru obiectivul O2, cum ar fi Pădurea aleatoare de arbori, Arborele de decizie, variante pentru Naive Bayes, diferite kerneluri pentru Mașina cu suport vectorial, Regresie Logistică, *Adaptive Boosting*, *Gradient Boosting Tree* etc., multiple tipuri de modele ansamblu, de pildă ansamblul cu vot majoritar, cu vot majoritar ponderat, ansambluri folosind algoritmul *Particle Swarm Optimization* și strategii de învățare profundă (de exemplu, ResNet50V2, ResNet101, VGG16, arhitecturi convoluționale). În plus, experimentele au luat în considerare multiple seturi de date (obiectivul O3), cum ar fi un set de date cu caracteristici lexicale și de tip *host-based* deja extrase și două seturi de date mai mari (adică D1 și D2) conținând doar linkul web și eticheta. Mai mult decât atât, caracteristicile alese au fost diverse (obiectivul O4), unele deja extrase din setul de date. Alte experimente au încorporat caracteristici colectate automat

din site-urile web. Un alt set de teste a folosit o formulă de embedding (adică TF-IDF), iar în final, o altă lucrare a luat în considerare codificarea datelor URL într-o imagine în nuanțe de gri sau color.

Efectuând experimente pe setul de date cu caracteristici deja extrase din linkuri web, *K-Nearest Neighbor* a îmbunătățit soluția din stadiul actual al literaturii de specialitate, iar RF a obținut cea mai bună precizie cu 87,6%. Mai mult, a existat o metodologie propusă pentru antrenarea și compararea algoritmilor. Aceasta a fost aplicată pentru *K-Nearest Neighbor*, Pădurea aleatoare de arbori și Arborele de decizie. Caracteristicile de tip *host-based* s-au dovedit a fi mai relevante comparativ cu atributele lexicale. Pe același set de date, a fost adăugată o comparație între alți algoritmi de învățare automată. Au fost construite modele ansamblu heterogene formate din trei clasificatori. Toate modelele au fost calibrate și toate combinațiile au fost încercate. Ansamblurile de clasificatori îmbunătățesc performanța modelelor individuale, ajungând la cel mult o precizie de 96,31% obținută cu ADA-XGB-SVM-RBF. Această soluție reușește să depășească unele soluții din literatura de specialitate prin modele precum Perceptronul Multistrat, Regresie Logistică și Naive Bayes Gaussian.

Studiile empirice asupra setului de date cu URL-uri, D1, au reușit să depășească alte soluții anterioare. Abordarea propusă consideră ansambluri heterogene formate din mai mulți clasificatori de învățare automată cu ponderile generate de algoritmul *Particle Swarm Optimization*. Soluția a ajuns la 97,05% acuratețe.

Pentru al doilea și cel mai mare set de date URL, D2, experimentele cu atributele selectate manual au obținut 94-95% acuratețe pe întregul set de date. Adăugând clasificarea generată de modelul de la OpenAI, s-au observat îmbunătățiri minore pentru majoritatea algoritmilor rulați și a ansamblurilor cu vot majoritar ponderat. În plus, privind setul de date D2, o soluție nouă a fost propusă prin transformarea URL-urilor în imagini color sau alb-negru. Apoi, o varietate de arhitecturi de învățare profundă au fost alese pentru experimente, cum ar fi Rețele Neuronale Convoluționale, mecanismul cu memorie de lungă și de scurtă durată (*Long Short Term Memory*), VGG16, ResNet50V2 și ResNet101. Cel mai bun model s-a dovedit a fi ResNet50V2 cu stratul *Long Short Term Memory*, atingând o performanță de 96,82% acuratețe. A obținut un scor mai mare comparativ cu alte rezultate din literatură.

În continuare, relația dintre linkurile web malițioase, clickbait și știrile false a fost investigată, începând cu detecția clickbait și a știrilor false pentru care s-au folosit algoritmi inteligenți și caracteristici extrase manual. Luând în considerare obiectivele propuse, colectarea eșantioanelor non-clickbait prin serverul proxy Squid bazat pe Hypertext Transfer Protocol (Protocol de transfer hipertext) (HTTP) și euristica propusă nu au fost potrivite pentru un scenariu real, astfel încât puține înregistrări au fost colectate în acest fel (obiectivul O5).

Prin urmare, la următorul experiment a fost inclus setul de date propus de *Clickbait Challenge*. Un alt obiectiv (O6) a fost utilizarea mai multor algoritmi de învățare automată (de exemplu, Pădurea aleatoare de arbori, Regresia Logistică, Naive Bayes și Mașina cu suport vectorial), care a fost realizat cu succes. Caracteristicile au fost extrase astfel încât să fie independente de limbă (obiectivul O7). Rezultatele obținute, și anume: acurateţe între 71% și 74%, precizie între 72% și 75%, recall în intervalul [68-71]% și scor $F_1$ între 70% și 73%, calculate pentru Regresia Logistică, Mașina cu suport vectorial și Pădurea aleatoare de arbori, sunt pertinente și comparabile cu alte soluții. Acestea încurajează mai multă cercetare în această direcție. Analiza făcută asupra celor mai semnificative atribute este confirmată de rezultate similare obținute în literatura științifică (obiectivul O8).

Luând în considerare detecția știrilor false, soluția propusă implică antrenarea unui model rulat pe o bază de date cu înregistrări în limba engleză și germană. Caracteristicile utilizate pentru modelare au fost considerate a fi independente de limba setului de date (obiectivul O7). Având obiectivul O6 în minte, mai mulți algoritmi inteligenți au fost comparați și cel mai bun (Pădurea aleatoare de arbori pentru modelul cu toate categoriile de caracteristici) a reușit să obțină o acurateţe de 75,38%. În plus, cele mai relevante caracteristici au fost clasificate și discutate în comparație cu rezultatele din alte lucrări din domeniu (obiectivul O8).

Luând în considerare relația dintre linkurile web malițioase, știrile false și mesajele clickbait sau titlurile, această asociere nu este foarte comună; este un scenariu de nișă, foarte dificil de capturat în viața reală. Cu toate acestea, au fost găsite câteva articole și opinii pentru această nișă de cercetare (obiectivul O9).

## 6.2 Direcții viitoare de cercetare

Ca direcție viitoare pentru continuarea și aprofundarea recenziei literaturii de specialitate privind detecția linkurilor web malițioase(obiectivul O1), cercetarea poate fi continuată și merită o analiză exhaustivă, incluzând statistici, dacă soluțiile pot fi implementate în mediu industrial, dacă sunt scalabile și cât de robuste sunt ele.

Considerând cercetarea viitoare privind domeniul detecției linkurilor web malițioase și obiectivul O2, acestea ar putea fi direcționată către modele de inteligență artificială explicabile, ajutând atât specialiștii, cât și utilizatorii să fie mai bine informați. Mai mult, explicabilitatea poate contribui la creșterea conștientizării tacticilor înșelătoare utilizate de atacatori. De asemenea, ar trebui elaborate mai multe tipuri de modele ansamblu și hibride, care să conțină clasificatori inteligenți și modele de învățare profundă. Ansamblurile dezvoltate cu algoritmi inspirați din natură ar putea fi o direcție

de investigat. Altă nișă de cercetare poate fi reprezentată de utilizarea modelelor mari de limbaj pentru diferite sarcini, cum ar fi clasificarea linkurilor, împreună cu detecția clickbait-ului și a știrilor false. Acestea pot fi folosite și pentru alte sarcini și pot îmbunătăți clasificarea realizată cu algoritmi inteligenți.

Implementarea la scară largă a soluțiilor propuse pot îmbunătăți acuratețea și robustețea modelului în timp (obiectivul O3). De asemenea, seturi de date de dimensiuni mari pot testa dacă detecția este scalabilă în mediul real și dacă soluția este suficient de robustă și precisă. În plus, sistemul de clasificare a linkurilor ar trebui testat pentru atacuri adversariale.

Pe de altă parte, ideea de a transforma URL-urile în imagini are potențial și ar putea fi investigată în continuare pentru a codifica mai bine informațiile din linkurile web. De exemplu, imaginea ar putea avea diferite regiuni unde informațiile specifice dintr-un URL vor fi codificate (de exemplu, protocol, adresă IP dacă este folosită, port, nume de domeniu etc.). Alte tipuri de caracteristici ar putea fi adăugate, mai ales caracteristici extrase automat folosind autoencodere, modele mari de limbaj și altele. În plus, colectarea URL-urilor poate fi uneori o direcție relevantă de investigație deoarece multe medii unde acestea se răspândesc sunt criptate, de exemplu, în cazul aplicației de mesagerie WhatsApp. Astfel, în aceste aplicații, interceptarea linkului web poate fi o idee nouă pentru analiză ulterioară. Acest lucru ar putea ajuta la avansarea rezultatelor noastre obținute pentru obiectivul O4.

Luând în considerare colectarea înregistrărilor de tip non-clickbait (obiectivul O5), o direcție viitoare ar putea fi dezvoltarea de euristici mai complexe sau chiar utilizarea metodelor nesupervizate pentru analiza traficului de rețea, urmărind ce site-uri sunt clickbait sau nu. Pentru îmbunătățirea obiectivului O6, se poate experimenta cu arhitecturi de învățare profundă și chiar modele mari de limbaj, pentru a face clasificarea pentru clickbait și știri false. În plus, pentru obiectivul O7, sistemul ar trebui să includă și alte tipuri de caracteristici, cum ar fi imaginile anexate în articolul de știri. De asemenea, atributele ar putea fi extrase automat și date mai departe algoritmului de clasificare. Caracteristicile create automat au avantajul de a fi independente de limbă. Obiectivul O8 poate fi îmbunătățit prin testarea algoritmilor pe date suplimentare, pe știri care tratează o mare varietate de subiecte. Astfel, se poate obține o perspectivă mai semnificativă asupra celor mai relevante indicii în detecția clickbait-ului și a știrilor false.

Având în vedere asocierea dintre clickbait, știri false și linkurile care răspândesc malware, aceasta poate fi investigată considerând o perspectivă interdisciplinară pentru a avansa cu obiectivul O9. Un sistem integrat care colectează și clasifică știri false, clickbait și linkuri ar putea fi dezvoltat și pus la dispoziție pe Internet. Altă idee interesantă de cercetare este considerarea clickbait-ului și a legăturilor web malițioase ca o problemă de regresie comparativ cu o problemă de clasificare.

# Appendix A

# Listă de abrevieri

**ADA** Adaptive Boosting

**ADA-XGB-SVM-rbf** Adaptive Boosting - Gradient Boosted Tree - Support Vector Machine with Radial Basis Function

**ASCII** American Standard Code for Information Interchange (Codul standard american pentru schimb de informații)

**CNN** Convolutional Neural Network (Rețea neuronală convoluțională)

**CSS** Cascading Style Sheets

**DT** Decision Tree (Arbore de decizie)

**GPT** Generative Pre-trained Transformer (Transformator generativ pre-antrenat)

**HTTP** Hypertext Transfer Protocol (Protocol de transfer hipertext)

**HTML** Hypertext Markup Language

**JS** JavaScript

**KNN** K-Nearest Neighbor

**LDA** Linear Discriminant Analysis

**LR** Logistic Regression (Regresie Logistică)

**LSTM** Long Short Term Memory (Mecanism de memorie de scurtă și de lungă durată)

**ML** Machine Learning (Învățare automată)

**MLP** Multi-layer Perceptron (Perceptron cu straturi multiple)

**NB** Naive Bayes

**NC** Nearest Centroid

**PAC** Passive Aggressive Classifier (Clasificator pasiv-agresiv)

**PSO** Particle Swarm Optimization

**RBF** Radial Basis Function

**RF** Random Forest (Pădure aleatoare de arbori)

**SGD** Stochastic Gradient Descent

**SVM** Support Vector Machine (Mașină cu suport vectorial)

**TF-IDF** Term Frequency and Inverse Document Frequency

**URL** Uniform Resource Locator

**WWW** World Wide Web

**XGB** Gradient Boosted Tree

# Bibliography

## Publicațiile noastre


[1] **Coste, Claudia Ioana**. "Controlling the click bait". In: *Proceedings of the International Student Conference StudMath-IT*. 2018, pp. 11–17.

[2] **Coste, Claudia Ioana**. "Online bad habits: fake news and clickbait". In: *Proceedings of the Student Interdisciplinary Conference The European Union and Global Order*. Cluj Universitary Press, 2019, pp. 179–193.

[3] **Coste, Claudia Ioana**. "Using Ensemble Models for Malicious Web Links Detection." In: *ICAART (3)*. 2024, pp. 657–664.

[4] **Coste, Claudia Ioana** and Andreica, Anca-Mirela and Chira, Camelia. "Malicious Web Links Detection Using Ensemble Models." In: *WEBIST*. 2023, pp. 268–275.

[5] **Coste, Claudia Ioana** and Bufnea, Darius. "Advances in clickbait and fake news detection using new language-independent strategies". In: *Journal of Communications Software and Systems* 17.3 (2021), pp. 270–280.

[6] **Coste, Claudia Ioana** and Bufnea, Darius and Niculescu, Virginia. "A new language independent strategy for clickbait detection". In: *2020 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. IEEE. 2020, pp. 1–6.

[7] **Coste, Claudia Ioana** and Chira, Camelia and Bufnea, Darius-Vasile. "Malicious Web Links Detection: A Survey". In: *SUBMITTED* -.- (2026).

[8] **Coste, Claudia-Ioana**. "Clickbait & Fake News - a cross language approach". Masters Thesis. Babeș-Bolyai University, Cluj-Napoca, Romania, 2021.

[9] **Coste, Claudia-Ioana**. "Contribuții privind identificarea legăturilor web de tip „clickbait" (eng., Contributions Regarding Clickbait Web Links)". Bachelor's Thesis. Babeș-Bolyai University, Cluj-Napoca, Romania, 2019.

[10] **Coste, Claudia-Ioana**. "Malicious Web Links Detection - A Comparative Analysis of Machine Learning Algorithms". In: *Studia Universitatis Babeș-Bolyai Informatica* 68.1 (2023), pp. 21–36. ISSN: 2065-9601. DOI: 10.24193/subbi.2023.1.02. URL: https://www.cs.ubbcluj.ro/~studia-i/journal/journal/article/view/86.

[11] **Coste, Claudia-Ioana**. "Malicious Web Links Detection Based on Image Processing and Deep Learning Models". In: *2024 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. IEEE. 2024, pp. 445–449. ISBN: 979-8-3315-0494-6/24. DOI: 10.1109/WI-IAT62293.2024.00071.

[12] **Coste, Claudia-Ioana.** and Kaisser, Thomas. "Malicious Web Links Detection with Large Language Models." In: *Lecture Notes in Business Information Processing* Accepted for publication.In press. (2026).

[13] Kaisser, Thomas. and **Coste, Claudia-Ioana.** "Using ChatGPT for Malicious Web Links Detection." In: *WEBIST*. 2024, pp. 425–432. ISBN: 978-989-758-718-4.

## Restul publicațiilor

[14] Husam Adas, Sachin Shetty, and Waled Tayib. "Scalable detection of web malware on smartphones". In: *2015 International Conference on Information and Communication Technology Research (ICTRC)*. IEEE. Abu Dhabi, UAE: IEEE, 2015, pp. 198–201.

[15] Alexa Internet, Inc. *The top 500 sites on the web*. Alexa-TopSites (https://www.alexa.com/topsites) (accesat la data Februarie 2, 2023). 2023.

[16] Veronica Anghel. *Why Romania just canceled its presidential election*. News article (https://www.journalofdemocracy.org/online-exclusive/why-romania-just-canceled-its-presidential-election) (accesat la data Feb 13, 2025). Dec. 2024.

[17] Dominic DiFranzo and Kristine Gloria-Garcia. "Filter bubbles and fake news". In: *XRDS: crossroads, the ACM magazine for students* 23.3 (2017), pp. 32–35.

[18] Birhanu Eshete, Adolfo Villafiorita, and Komminist Weldemariam. "Malicious website detection: Effectiveness and efficiency issues". In: *2011 First SysSec Workshop*. IEEE. Amsterdam, The Netherlands: IEEE, 2011, pp. 123–126.

[19] Jessica Valasek Estenssoro. Cybersecurity Threats and Trends & Malware Statistics 2024 (https://www.avg.com/en/signal/malware-statistics) (accesat la data Nov 24, 2024). Nov. 2024.

[20] Kaggle Inc. *Kaggle*. Kaggle (https://www.kaggle.com/) (accesat la data Martie 20, 2025). 2022.

[21] GTKlondike. *Machine-Learning-for-Security-Analysts*. Dataset website (https://github.com/NetsecExplained/Machine-Learning-for-Security-Analysts) (accesat la data Iunie 22, 2024). June 2019.

[22] Davide Mancino and Frederica Fragapane. Digital Literacy in the EU (https://data.europa.eu/en/publications/datastories/digital-literacy-eu-overview) (accesat la data Nov 24, 2024). Dec. 2023.

[23] Oxford University Press. *Clickbait noun - definition, pictures, pronunciation and usage notes*. Clickbait definition (https://www.oxfordlearnersdictionaries.com/definition/english/clickbait?q=clickbait) (accesat la data Dec 20, 2024). Dec. 2024.

[24] F. et al. Pedregosa. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[25] Lexie Pelchen and Samantha Allen. *Internet usage statistics in 2024*. Forbes - Internet Statistics (https://www.forbes.com/home-improvement/internet/internet-statistics/) (accesat la data Nov 24, 2024). Nov. 2024.

[26] Ani Petrosyan. *Internet and social media users in the world 2024*. Statista article (https://www.statista.com/statistics/617136/digital-population-worldwide/) (accesat la data Nov 12, 2024). Nov. 2024.

[27] Kai Shu et al. "Fake News Detection on Social Media: A Data Mining Perspective". In: *SIGKDD Explor. Newsl.* 19.1 (Sept. 2017), pp. 22–36. ISSN: 1931-0145. DOI: 10.1145/3137597.3137600.

[28] Manu Siddhartha. *Malicious URLs dataset*. Kaggle - Malicious URLs dataset (https://www.kaggle.com/datasets/sid321axn/malicious-urls-dataset) (accesat la data Mai 2, 2024). July 2021.

[29] Damian Tambini. "Fake news: public policy responses". In: *London School of Economics and Political Science* (2017).

[30] Christian Urcuqui et al. "Machine Learning Classifiers to Detect Malicious Websites." In: *SSN*. Spain: SSN, 2017, pp. 14–17.

[31] *World Population Clock: 8.2 Billion People (LIVE, 2024) - Worldometer*. Worldmeter website (https://www.worldometers.info/world-population/) (accesat la data Nov 12, 2024). Nov. 2024.