

Raport 3

Proiect:

**SCAN-NEWS: Sistem inteligent pentru detecția și atenuarea răspândirii
dezinformării și știrilor false în rețelele sociale**

- Iunie 2025 -

Director proiect: Conf. Habil. Dr. Ing. Elena-Simona APOSTOL

Tranziția către era digitală a dus la o schimbare radicală a peisajului media, cu accent pe personalizarea conținutului și interacțiunea utilizatorilor. Pe măsură ce mass-media devine tot mai digitalizată, au apărut noi modele jurnaliste pentru diseminarea informațiilor. Aceste modele au transformat semnificativ modul în care societatea consumă știri și informații. În cursa de a rămâne înaintea concurenților, jurnaliștii sacrifică uneori standardele jurnalismului tradițional, prioritizând viteza și având ca scop "să devină virali" prin generarea rapidă de vizualizări, urmărind o creștere a aprecierilor, comentariilor și distribuțiilor. Această evoluție a diminuat importanța jurnalismului de investigație și a facilitat răspândirea rapidă a dezinformării. Cercetările actuale în acest domeniu se limitează adesea la analiza conținutului mesajelor, ignorând factorii sociali, culturali și politici care influențează receptarea și interpretarea acestora.

Raportul de față prezintă o analiză detaliată a activităților de cercetare și diseminare derulate în cadrul proiectului SCAN-NEWS, în intervalul Ianuarie – Iunie 2025.

1. Activități de cercetare

Proiectul SCAN-NEWS are ca prin focus analiza diferitelor tipuri de dezinformări prezente în rețelele sociale, precum:

- Știri false (en. *fake news*): reprezintă informații false prezentate ca fiind adevărate
- Dezinformare parțială: constă în informații adevărate, dar prezentate într-un mod înșelător
- Satire și umor confundate cu realitatea: sunt știri sau postări satirice sau umoristice care sunt interpretate greșit ca fiind adevărate
- Conspirații: constă în teorii nefondate care explică evenimente sau fenomene complexe prin intermediul unor explicații simple, dar false
- Dezinformare neintenționată (en. *misinformation*): sunt informații incorecte, dar răspândite fără intenție de a înșela, ci din simplă neștiință sau eroare
- Dezinformare intenționată (en. *disinformation*): reprezintă răspândirea deliberată de informații false sau distorsionate cu scopul de a manipula sau înșela
- Etc.

Obiectivul principal al proiectului SCAN-NEWS se concentrează pe două direcții majore, cercetarea și dezvoltarea de:

- [**O₁**] noi modele de detecție de tip ansamblu care să integreze, pe lângă contextul cuvintelor extras cu soluții bazare pe Transformers, informații despre polaritatea și tematica știrilor, precum și graful de difuzie al știrii în rețelele sociale, pentru o detecție mai precisă;
- [**O₂**] noi strategii de imunizare a rețelelor sociale care au ca scop stoparea sau limitarea răspândirii online a informațiilor false, contribuind la un mediu digital mai sigur.

Raportul 3 se focusează mai mult pe primul obiectiv (**O₁**). În cadrul direcției **O₁**, ne propunem să creăm noi tipuri de reprezentare a informațiilor pentru a îmbunătăți performanțele modelelor de detecție a diferitelor forme de dezinformare din rețelele sociale. Aceste reprezentări vectoriale vor include atât informațiile postate de utilizatori (numite în continuare context local), cât și detalii despre rețeaua socială (ex. prieteni,

urmăritori), reacțiile utilizatorilor, subiectele abordate și polaritatea acestora (definite de noi ca context global).

		Diagrama GANTT a proiectului																				
WP	Luna	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
WP1	Activități de management																					
S1.1	Management administrativ				D1a					D1b					D1c							D1d
S1.2	Management tehnic									D1e						D1f						
WP2	Analiza cerințelor și soluțiilor similare existente																					
S2.1	Analiza critică a soluțiilor existente				D2a																	
S2.2	Cerințe funcționale și non-funcționale																					
S2.5	Cerințe tehnologice					D2b																
WP3	Colectarea și adnotarea datelor																					
S3.1	Analiza seturilor de date disponibile online					D3a																
S3.2	Colectarea de noi date din mediile sociale								D3b													
S3.3	Preprocesare și extragerea caracteristicilor									D3c	M3a											
WP4	Proiectarea și implementarea sistemului																					
S4.1	Proiectarea arhitecturii																					
S4.2	Implementarea algoritmilor de Învățare Automată													D4a								
S4.3	Sciarea codului modulelor arhitecturii																	D4b				
S4.4	Integrarea modulelor																		D4c	M4a		
S4.5	Proiectarea interfeței cu utilizatorii																		D4d			
WP5	Evaluare și validare																					
S5.1	Definirea scenariilor de test															D5a						
S5.2	Testare experimentală a modulelor și sistemului														D5b	D5c	D5d					M5a
S5.3	Optimizări ale sistemului																				D5e	
WP6	Diseminare și exploatare																					
S6.1	Diseminarea rezultatelor								D6a					D6b		D6c			D6d			
S6.2	Exploatare sistem														M6a							M6b

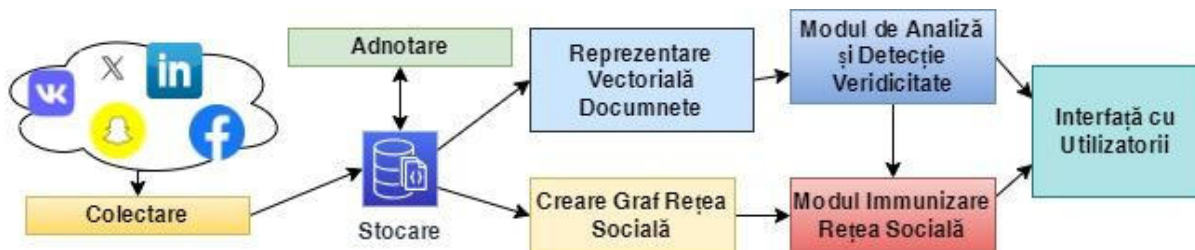
Figură 1. Diagrama GANTT

Conform calendarului de activități prezentat în propunerea de proiect și ilustrat în *Figură 1. Diagrama GANTT*, principalele activități de cercetare derulate în cadrul etapei 3 a proiectului SCAN-NEWS sunt următoarele:

- **(WP3) Colectarea și adnotarea datelor:** Am continuat activitățile de colectare, stocare și adnotare a datelor din platformele sociale, concentrându-ne în prezent pe platforma X (Twitter). Am adunat date atât în engleză, cât și în română, cu o preponderență a celor în limba engleză. Conform metodologiei stabilite anterior, am folosit același procesul de adnotare prezentat în rapoartele anterioare. Am dezvoltat o platformă web accesibilă adnotatorilor, unde aceștia primesc postări aleatorii și le etichetează conform claselor disponibile. Fiecare postare este adnotată de minimum trei adnotatori.
- **(WP3) Preprocesarea și extragerea caracteristicilor:** Am folosit tehnici clasice de preprocesare precum eliminarea adreselor URL, a spațiilor duble și a semnelor de punctuație. Am realizat și o analiză mai detaliată a caracteristicilor textuale ce a luat în considerare și marcatori de discurs folosiți în seturile de date. Marcatorii de discurs, precum "deci", "adică" sau "ei bine", joacă un rol important în structurarea discursului și în transmiterea nuanțelor de sens, iar înțelegerea lor poate oferi indicii valoroase despre stilul și intenția textului. Importanța acestei analize nu se limitează doar la contextul specific al proiectului nostru. Astfel, am extins cercetarea privind marcatorii de discurs și am publicat o analiză mai generală în *Wiley International Journal of Applied Linguistics*, acoperind atât limba română, cât și alte nouă limbi.

- **(WP4) Algoritm de detecție:** În cadrul acestei acțiuni s-a propus o nouă arhitectură de tip ansamblu ce folosește algoritmi de Deep Learning și îmbunătățită cu informații despre rețea, pentru detecția diferitelor tipuri de manipulări din rețelele sociale.
- **(WP5) Evaluare și validare:** În această etapă au fost definite mai multe scenarii de test pentru evaluarea modulelor implementate până la acest moment. O parte din rezultate au fost deja diseminate la diferite conferințe și jurnale.
- **(WP6) Diseminarea și exploatare:** Pentru această etapă, au fost publicate 2 articole de jurnal, ambele cotate ISI Q1 în *Lingvistică*, respectiv în *Informatică și Inteligență Artificială* (mai multe detalii în *Secțiunea 2. Activități de publicare*)

Arhitectura soluției finale este prezentată în *Figură 2. Arhitectura generală SCAN-NEWS*. Până la finalizarea raportului 3, am realizat sarcini de proiectare, implementare și testare pentru toate modulele. În cele ce urmează, vom detalia implementarea și rezultatele obținute pentru modulele dezvoltate și testate în etapa 3. Menționăm că acest raport include doar rezultatele deja publicate.



Figură 2. Arhitectura generală SCAN-NEWS

Analiza și Detecția Veridicității Informațiilor postate în Mediile Sociale

În această secțiune, vom detalia îmbunătățirile aduse algoritmilor de analiză și detecție a veridicității în cadrul proiectului. Ne vom concentra exclusiv pe noutățile introduse față de Rapoartele 1 și 2.

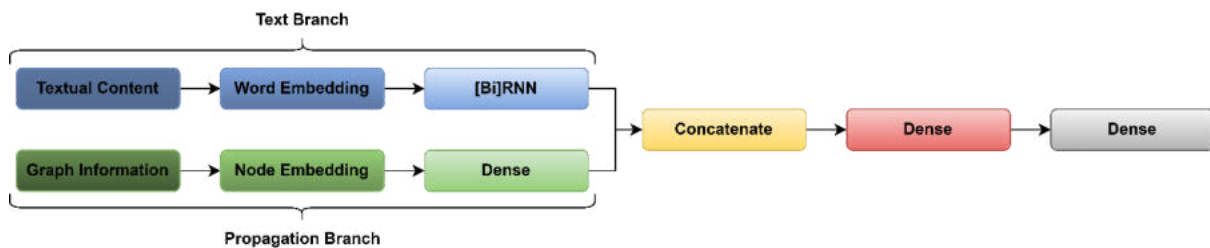
Cercetările actuale legate de detecția veridicității informațiilor postate în mediile sociale explorează modele de detecție a știrilor false, antrenate fie prin învățare automată tradițională, fie prin învățare profundă (deep learning), pe seturi mari de articole de știri. Câteva noi abordări integrează și metadate suplimentare, cum ar fi sursele de citare, autori verificabili, paragrafe scurte, narațiunea la persoana a treia și alți factori care pot îmbunătăți precizia modelelor de detecție a știrilor false. Cu toate acestea, antrenarea pe astfel de date poate avea dezavantaje, mai ales când modelul întâlnește date noi care nu respectă aceleași convenții de scriere. De exemplu, conținutul din social media este de obicei mai personal și informal.

Acesta este motivul pentru care integrarea de informații suplimentare este crucială în dezvoltarea sistemelor de detecție a știrilor false pentru social media.

Având în vedere aceste aspecte, în cadrul proiectului am propus o nouă arhitectură de rețea neuronală profundă de tip ansamblu îmbunătățită cu informații despre rețea, pentru detecția știrilor false. Am denumit această nouă arhitectură GETAE: **Graph Information Enhanced Deep Neural Network Ensemble Architecture for Fake News Detection**.

GETAE utilizează informațiile din rețelele sociale stocate sub formă de graf, creând reprezentări avansate ale informației prin folosirea de node embeddings. Aceste reprezentări, combinate cu conținutul textual, sunt apoi introduse în rețele neuronale de tip feed-forward sau recurente bidirecționale. Acesta permite modelului nostru să valorifice atât datele contextuale, cât și pe cele specifice rețelei sociale.

GETAE este alcătuit din două componente principale: Ramura Textuală (en. *Text Branch*) și Ramura de Propagare (en. *Propagation Branch*), ilustrate în Figura 3.



Figură 3. GETAE - arhitectură avansată pentru detecția veridicității informației din mediile sociale

Ramura Textuală a arhitecturii GETAE (a se vedea *Text Branch* din Figura 3) este compusă din trei straturi:

- Stratul de intrare (Conținutul Textual),
- Stratul de reprezentare a cuvintelor (Word Embedding), și
- Un strat ascuns ([Bi]RNN)

Stratul de intrare (Conținut Textual) necesită două intrări:

- matricea document-token $D \in N^{n \times k}$ și
- matricea de reprezentare a cuvintelor $W \in R^{m \times s}$

Al doilea strat este utilizat pentru a asocia token-urile din documentul D cu încadrările de cuvinte corespunzătoare din W.

Stratul ascuns ([Bi]RNN) utilizează fie unități de Rețele Neurale Recurente Unidirecționale (RNN), fie unități de Rețele Recurente Bidirecționale (BiRNN). În implementarea GETAE pentru testarea prin ablație, RNN-ul este înlocuit fie cu unități standard de Rețele Neurale Recurente (RNN), fie cu unități de Memorie pe Termen Lung (LSTM) sau cu unități Recurente cu Poartă (GRU).

Ieșirea stratului ascuns este o nouă reprezentare a cuvintelor, numită de noi în engleză *Text Content Embedding*. Aceasta reprezentare combină multiple caracteristici textuale, cum ar fi cele lexicale (ex. caracteristici la nivel de caracter și cuvânt) și sintactice (ex. caracteristici la nivel de propoziție).

Ramura de Propagare din arhitecturii GETAE este compusă din trei straturi:

- Stratul de intrare (Informațiile de tip Graf),
- Stratul de reprezentare a informației bazate pe noduri (en. *Node Embedding*), și

- Un strat ascuns (de tip Dense).

Stratul de intrare necesită două intrări:

- lista de noduri V și
- matricea de încadrare a nodurilor $N \in R^{|V| \times s}$

Stratul bazat pe Node Embedding este folosit pentru a asocia nodurile din V cu reprezentările corespunzătoare ale nodurilor din N . Stratul ascuns utilizează aceste perechi pentru a obține o nouă reprezentare denumită de noi Propagation Embedding. Acest strat ascuns conține unități Dense care folosesc funcția de activare *ReLU* pentru a crea noua reprezentare (en. *Propagation Embedding*). Această reprezentare a fost propusă pentru a codifica difuzia informației, adică răspândirea informațiilor de la un nod la următorii săi.

În continuare, GETAE unește rezultatele Ramurilor Textuală și de Propagare într-un singur tensor. Acesta este apoi transmis unui strat Dense utilizat pentru clasificare. Prin această concatenare, am creat o nouă reprezentare a conținutului îmbunătățită prin informații de propagare (en. *Propagation-Enhanced Content Embedding*). Aceasta ia în considerare atât conținutul textual, cât și modul în care informația se propagă în graful rețelei sociale, combinând astfel Reprezentarea Conținutului Textual cu Reprezentarea Informației de Propagare. Această nouă reprezentare este apoi procesată de un strat Dense care extrage caracteristici contextuale ascunse și generează o nouă reprezentare vectorială, folosind funcția de activare *ReLU*.

Noua reprezentare vectorială este, în cele din urmă, pasată printr-un ultim strat Dense pentru clasificare. Acesta utilizează funcția de activare *softmax* pentru a determina veridicitatea unei postări.

Pentru antrenarea și validarea experimentală a arhitecturii GETAE, am utilizat seturile de date de pe Twitter, conținând următoarele câmpuri:

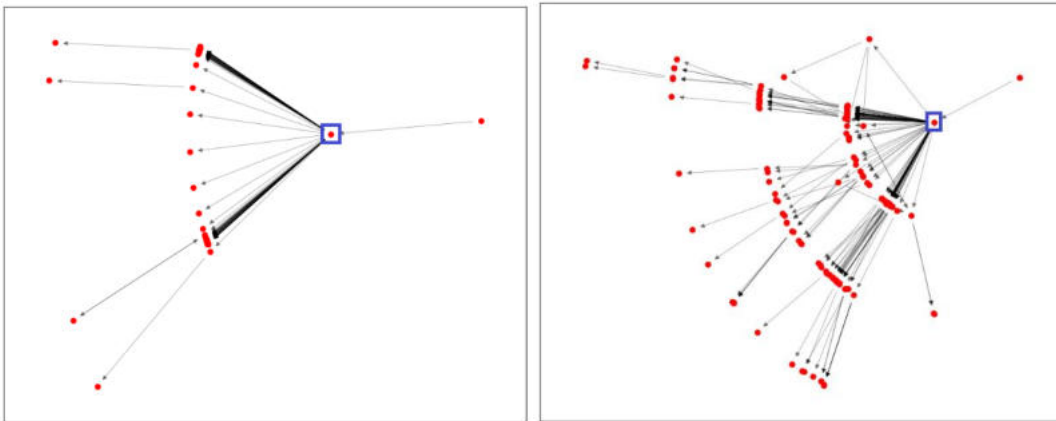
- tweet-urile sursă cu ID-ul tweet-ului și conținutul său textual,
- eticheta, care atribuie o clasă fiecărui ID de tweet,
- arborele de propagare pentru fiecare tweet.

Seturi de date sunt etichetate folosind patru categorii: fals, adevărat, non-zvon și neverificat. Din analiza seturilor de date am observat că lungimea cea mai frecventă este de aproximativ 80-100 de caractere iar majoritatea tweet-urilor se încadrează în intervalul de 10-15 cuvinte, cu un număr maxim de cuvinte sub 30. Aceasta subliniază caracteristica tweet-urilor ca fiind texte scurte, cu puține cuvinte și un număr redus de caractere.

Pentru evaluare am folosit ca metrici: Acuratețea, Precizia, Amintirea (Recall) și Scorul F1.

Pentru seturile de date folosite, fiecare arbore de propagare este codificat printr-o listă de muchii (legături) de la nodurile părinte la cele copil. Relația dintre un nod copil și un nod părinte reprezintă retweet-uri și mențiuni ale tweet-ului sursă, dar numai pentru vecinii nodului rădăcină.

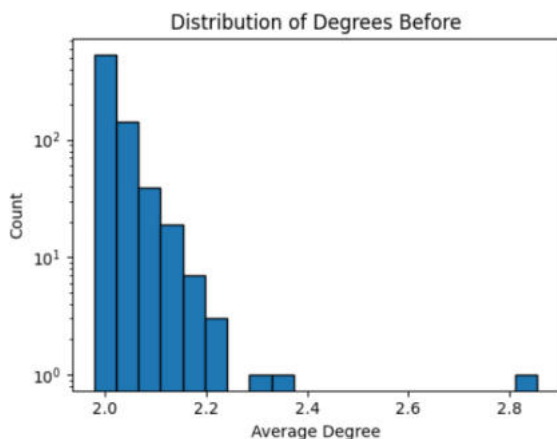
Figura 4 prezintă câteva exemple de arbori de propagare pentru o mai bună vizualizare a modului în care informația se răspândește. Acești arbori de propagare sunt folosiți pentru a construi reprezentările informației de propagare cu Node2Vec și DeepWalk, pe care GETAE le utilizează pentru a crea noua reprezentare propusă de noi (en. *Propagation-Enhanced Content Embedding*).



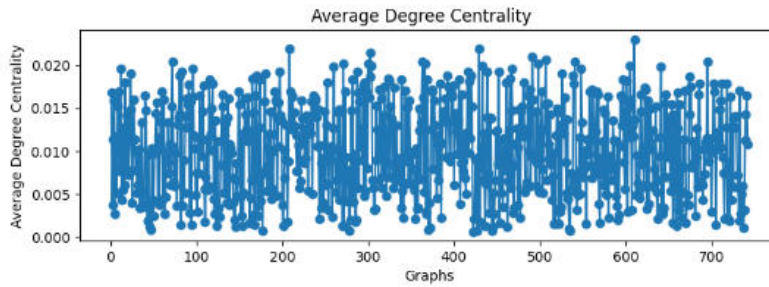
Figură 4. Arbori de propagare

Grafurile utilizate pentru reprezentarea informației de propagare sunt orientate, respectând direcția specificată în setul de date. Fiecare rând din fișierul de muchii (edges) reprezintă o conexiune de la un nod părinte la un nod copil. Nodurile sunt reprezentate sub formă de tupluri cu trei componente: ID-ul utilizatorului de Twitter ($user_{id}$), ID-ul tweet-ului ($tweet_{id}$), și timpul în minute scurs de la postarea primului nod (nodul sursă).

Figura 5 ilustrează distribuțiile gradelor grafurilor sub formă de histograme, utilizând o scară logaritmică. Pentru primul set de date Twitter, care include un total de 742 de grafuri, am calculat gradul mediu pentru fiecare dintre acestea. Observăm că majoritatea grafurilor prezintă o distribuție a gradelor în jurul valorii de 2, unele chiar depășind 2.8.



Figură 5. Distribuția gradelor grafurilor din setul de date



Figură 6. Centralitatea medie a gradului pentru nodurile din rețea

Figura 6 prezintă centralitatea medie a gradului pentru nodurile din rețea și este utilizată pentru a evalua riscul imediat de "infecție" al unui nod prin numărarea vecinilor săi.

Pentru antrenare și validare, am utilizat o rată de divizare de 80%-20% a seturilor de antrenare și testare. Am antrenat modele folosite în arhitectura GETAE timp de 30 de epoci. Excepție au fost modelele care au folosit reprezentarea de cuvinte Word2Vec Skip-Gram, acestea fiind antrenate doar pentru 8 epoci pentru a preveni supraînvățarea (en. *overfitting*). Pe parcursul antrenării, am aplicat o rată de învățare de 0.001. Procesul de antrenare pentru toate modelele a fost realizat pe Google Colab, fără accelerare hardware.

Toate experimentele au folosit validarea încrucișată k-fold cu k=10, menținând raportul de 80%-20% pentru seturile de antrenare și testare.

Prezentăm mai jos o parte din rezultatele obținute pe setul de testare după antrenarea modelelor GETAE pe seturile de date Twitter (curent platforma X). Am utilizat trei tipuri de reprezentare vectorială a cuvintelor (Word2Vec Skip-Gram, BERT și BERTweet) și două tipuri de reprezentare a nodurilor din rețeaua socială (Node2Vec și DeepWalk).

Word Embedding	Network Embedding	Text Branch	Propagation Branch	RNN Layer				BiRNN Layer			
				Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
WORD2VEC	N/A	✓	N/A	0.749 ± 0.039	0.764 ± 0.036	0.750 ± 0.042	0.745 ± 0.043	0.754 ± 0.027	0.763 ± 0.027	0.754 ± 0.028	0.751 ± 0.027
	NODE2VEC	✓	✓	0.651 ± 0.028	0.683 ± 0.025	0.660 ± 0.027	0.642 ± 0.041	0.664 ± 0.022	0.698 ± 0.042	0.675 ± 0.025	0.656 ± 0.023
	DEEPWALK	✓	✓	0.705 ± 0.025	0.720 ± 0.023	0.711 ± 0.023	0.703 ± 0.025	0.703 ± 0.043	0.729 ± 0.028	0.711 ± 0.039	0.698 ± 0.049
BERT	N/A	✓	N/A	0.790 ± 0.036	0.791 ± 0.034	0.788 ± 0.035	0.784 ± 0.037	0.797 ± 0.033	0.800 ± 0.034	0.797 ± 0.033	0.794 ± 0.034
	NODE2VEC	✓	✓	0.766 ± 0.070	0.800 ± 0.035	0.769 ± 0.066	0.755 ± 0.090	0.800 ± 0.028	0.810 ± 0.024	0.801 ± 0.029	0.797 ± 0.029
	DEEPWALK	✓	✓	0.736 ± 0.063	0.773 ± 0.044	0.738 ± 0.060	0.725 ± 0.071	0.795 ± 0.031	0.802 ± 0.032	0.796 ± 0.030	0.791 ± 0.031
BERTWEET	N/A	✓	N/A	0.735 ± 0.084	0.780 ± 0.024	0.736 ± 0.086	0.714 ± 0.128	0.792 ± 0.059	0.824 ± 0.026	0.797 ± 0.053	0.787 ± 0.064
	NODE2VEC	✓	✓	0.701 ± 0.097	0.773 ± 0.042	0.698 ± 0.095	0.664 ± 0.139	0.790 ± 0.050	0.812 ± 0.032	0.790 ± 0.044	0.783 ± 0.054
	DEEPWALK	✓	✓	0.736 ± 0.075	0.772 ± 0.039	0.725 ± 0.084	0.711 ± 0.113	0.827 ± 0.041	0.839 ± 0.031	0.827 ± 0.044	0.823 ± 0.046
Word Embedding	Network Embedding	Text Branch	Propagation Branch	GRU Layer				BiGRU Layer			
				Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
WORD2VEC	N/A	✓	N/A	0.481 ± 0.020	0.240 ± 0.010	0.500 ± 0.000	0.324 ± 0.008	0.631 ± 0.043	0.696 ± 0.047	0.636 ± 0.044	0.603 ± 0.066
	NODE2VEC	✓	✓	0.485 ± 0.027	0.351 ± 0.150	0.510 ± 0.017	0.371 ± 0.085	0.614 ± 0.029	0.702 ± 0.053	0.627 ± 0.024	0.584 ± 0.052
	DEEPWALK	✓	✓	0.485 ± 0.033	0.491 ± 0.034	0.492 ± 0.033	0.469 ± 0.044	0.625 ± 0.027	0.656 ± 0.045	0.631 ± 0.034	0.614 ± 0.031
BERT	N/A	✓	N/A	0.764 ± 0.031	0.772 ± 0.027	0.771 ± 0.025	0.761 ± 0.029	0.780 ± 0.018	0.784 ± 0.018	0.783 ± 0.017	0.779 ± 0.018
	NODE2VEC	✓	✓	0.757 ± 0.044	0.765 ± 0.041	0.760 ± 0.046	0.755 ± 0.045	0.768 ± 0.030	0.789 ± 0.025	0.766 ± 0.027	0.764 ± 0.029
	DEEPWALK	✓	✓	0.751 ± 0.036	0.751 ± 0.036	0.750 ± 0.036	0.749 ± 0.035	0.776 ± 0.015	0.776 ± 0.017	0.775 ± 0.014	0.773 ± 0.016
BERTWEET	N/A	✓	N/A	0.686 ± 0.044	0.734 ± 0.038	0.688 ± 0.040	0.667 ± 0.060	0.742 ± 0.038	0.753 ± 0.030	0.746 ± 0.036	0.742 ± 0.038
	NODE2VEC	✓	✓	0.694 ± 0.072	0.757 ± 0.038	0.695 ± 0.058	0.670 ± 0.091	0.731 ± 0.023	0.752 ± 0.033	0.735 ± 0.019	0.727 ± 0.021
	DEEPWALK	✓	✓	0.712 ± 0.044	0.745 ± 0.034	0.713 ± 0.049	0.701 ± 0.057	0.748 ± 0.029	0.754 ± 0.029	0.750 ± 0.028	0.747 ± 0.028
Word Embedding	Network Embedding	Text Branch	Propagation Branch	LSTM				BiLSTM Layer			
				Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
WORD2VEC	N/A	✓	N/A	0.667 ± 0.051	0.732 ± 0.046	0.668 ± 0.045	0.643 ± 0.066	0.669 ± 0.038	0.712 ± 0.042	0.673 ± 0.037	0.653 ± 0.045
	NODE2VEC	✓	✓	0.623 ± 0.040	0.649 ± 0.053	0.630 ± 0.040	0.617 ± 0.042	0.590 ± 0.037	0.674 ± 0.063	0.605 ± 0.039	0.556 ± 0.056
	DEEPWALK	✓	✓	0.642 ± 0.028	0.679 ± 0.034	0.653 ± 0.028	0.632 ± 0.030	0.647 ± 0.023	0.688 ± 0.035	0.657 ± 0.021	0.636 ± 0.036
BERT	N/A	✓	N/A	0.756 ± 0.049	0.788 ± 0.019	0.758 ± 0.043	0.746 ± 0.054	0.799 ± 0.030	0.805 ± 0.025	0.796 ± 0.034	0.794 ± 0.036
	NODE2VEC	✓	✓	0.783 ± 0.045	0.805 ± 0.031	0.775 ± 0.049	0.773 ± 0.053	0.827 ± 0.042	0.831 ± 0.035	0.827 ± 0.040	0.825 ± 0.042
	DEEPWALK	✓	✓	0.764 ± 0.050	0.782 ± 0.035	0.760 ± 0.050	0.757 ± 0.059	0.814 ± 0.043	0.818 ± 0.041	0.815 ± 0.042	0.812 ± 0.044
BERTWEET	N/A	✓	N/A	0.651 ± 0.105	0.758 ± 0.040	0.666 ± 0.096	0.611 ± 0.139	0.748 ± 0.049	0.777 ± 0.029	0.752 ± 0.042	0.741 ± 0.053
	NODE2VEC	✓	✓	0.708 ± 0.078	0.783 ± 0.028	0.702 ± 0.078	0.675 ± 0.117	0.758 ± 0.047	0.787 ± 0.035	0.759 ± 0.049	0.749 ± 0.058
	DEEPWALK	✓	✓	0.732 ± 0.074	0.762 ± 0.037	0.735 ± 0.062	0.720 ± 0.093	0.770 ± 0.083	0.815 ± 0.029	0.774 ± 0.077	0.757 ± 0.103

Acest tabel prezintă, de asemenea, două tipuri de testare pentru arhitectura GETAE:

- adăugarea sau eliminarea Ramurii de Propagare
- modificarea stratului RNN pentru Ramura Textuală

Ramura Textuală utilizează șase straturi diferite de rețele neuronale recurente pentru testarea de tip ablație: RNN, BiRNN, GRU, BiGRU, LSTM și BiLSTM.

Pentru acest set de date, cea mai bună configurație a GETAE, conform metricilor, este o egalitate între (BERT, Node2Vec, BiLSTM) și (BERTweet, DeepWalk, BiRNN). Comparând aceste două modele cu alte abordări de ultimă generație, am ales configurația (BERT, Node2Vec, BiLSTM) datorită unui scor F1 mai mare. De asemenea, BERTweet merită menționat pentru rezultatele sale promițătoare, în special atunci când utilizează încadrarea de noduri DeepWalk și BiRNN ca strat recurent.

Experimente realizate și prezentate de noi în această etapă pentru arhitectura GETAE demonstrează că procesul de detectare a știrilor false este îmbunătățit semnificativ prin valorificarea conținutului utilizatorilor și a interacțiunii lor în cadrul unei rețele sociale.

Ca și concluzie finală, soluția propusă de noi, GETAE, este o arhitectură de tip ansamblu promițătoare pentru antrenarea modelelor de detecție a știrilor false, luând în considerare propagarea informației în rețea. Prin analiza modului în care nodurile dăunătoare "infectează" rețeaua, modelele antrenate cu GETAE depășesc performanțele modelelor existente de ultimă generație.

Mai mult, GETAE reușește să detecteze mai bine știrile false prin combinarea conținutului textual cu propagarea acestuia în social media, valorificând interacțiunile utilizatorilor. Astfel, prin integrarea dinamicii sociale dintre utilizatori și a modului în care aceștia interacționează cu conținutul, GETAE subliniază și modul în care actorii malițioși reușesc să utilizeze abuziv informația și cum pot profita de prejudecățile utilizatorilor pentru a-i manipula.

Rezultatele prezentate în această secțiune au fost publicate în jurnalul Elsevier *Expert Systems With Applications*.

2. Activități de publicare

În cadrul proiectului SCAN-NEWS au fost publicate, cu afilierea AOȘR 8 articole. Cinci dintre cele opt articole sunt de jurnal și sunt cotate Q1 ISI conform Journal Citation Reports (JCR) pe 2023 respectiv 2024. Progresul actual depășește programul de diseminare stabilit în propunerea de proiect (Figură 1. Diagrama GANTT).

Dintre cele opt publicații, două articole sunt raportate în iunie 2025, restul au fost raportate în primul an al proiectului (Rapoartele 1 și 2).

Lista publicațiilor pentru proiectul SCAN-NEWS este redată mai jos.

Publicații raportate în Iunie 2025

- **Elena-Simona Apostol**, Ciprian-Octavian Truică, Mariana Damova, Purificação Silvano, Giedre Valunaite Oleškevicienė, Chaya Liebeskind, Dimitar Trajanov, Anna Baczkowska, Emma Angela Montecchiari, Christian Chiarcos. *Multiword Discourse Markers Across Languages: A Linguistic and Computational Perspective*, International Journal of Applied Linguistics, Wiley, ISSN 0802-6106, Aceptat spre publicare în Aprilie 2025. DOI: [10.1111/ijal.12755](https://doi.org/10.1111/ijal.12755) (Jurnal cotate ISI Q1 în Lingvistică, F.I. = 1.7 conform JCR 2024)
- Ciprian-Octavian Truică, **Elena-Simona Apostol**, Marius Marogel, Adrian Paschke. *GETAE: Graph Information Enhanced Deep Neural Network Ensemble*. Expert Systems with Applications, vol. 275, ISSN 0957-4174, Mai 2025, DOI: [10.1016/j.eswa.2025.126984](https://doi.org/10.1016/j.eswa.2025.126984) (Jurnal cotate ISI Q1 în Computer Science și Inteligență Artificială, F.I. = 7.5 conform JCR 2024)

Publicații raportate în Decembrie 2024

- Alexandru Petrescu, Ciprian-Octavian Truică, **Elena-Simona Apostol**, Adrian Paschke. *EDSA-Ensemble: an Event Detection Sentiment Analysis Ensemble Architecture*. IEEE Transactions on Affective Computing, 16(2): 555-572, ISSN: 1949-3045, April 2025, Disponibil online: August 2024, Publicat: Aprilie 2025, DOI: [10.1109/TAFFC.2024.3434355](https://doi.org/10.1109/TAFFC.2024.3434355) (Jurnal cotate ISI Q1, F.I. = 9.6 conform JCR 2023)
- **Elena-Simona Apostol**, Ciprian-Octavian Truică, Adrian Paschke. *ContCommRTD: A Distributed Content-based Misinformation-aware Community Detection System for Real-Time Disaster Reporting*. Transactions on Knowledge and Data Engineering Journal, November 2024. DOI: [10.1109/TKDE.2024.3417232](https://doi.org/10.1109/TKDE.2024.3417232) (Jurnal cotate ISI Q1, F.I. = 8.9 conform JCR 2023)
- Ciprian-Octavian Truică, Ana-Teodora Constantinescu, **Elena-Simona Apostol**. *StopHC: A Harmful Content Detection and Mitigation Architecture for Social Media Platforms*. IEEE International Conference on Intelligent Computer Communication and Processing (ICCP 2024), Octombrie 2024 (Conferință ISI cotate Național în CORE)

Publicații raportate în Iulie 2024

- **Elena-Simona Apostol**, Özgür Coban, Ciprian-Octavian Truică. *CONTAIN: A community-based algorithm for network immunization*. Engineering Science and Technology, an International Journal. Elsevier. 55:1-10(101728), ISSN 2215-0986, Iulie 2024. DOI: [10.1016/j.jestch.2024.101728](https://doi.org/10.1016/j.jestch.2024.101728) (Jurnal cotatec ISI Q1, F.I. = 5.1 conform JCR 2023)
- **Elena-Simona Apostol**, Adrian-Cosmin Cojocaru, Ciprian-Octavian Truică. *Large-Scale Graphs Community Detection using Spark GraphFrames*. The 23rd International Symposium on Parallel and Distributed Computing (ISPDC 2024). Iulie 2024 (Conferință ISI cotatec C în CORE)
- Alexandru Petrescu, Ciprian-Octavian Truică, **Elena-Simona Apostol**. *Language-based Mixture of Transformers for EXIST2024*. Conference and Labs of the Evaluation Forum (CLEF2024). Publicată în Lecture Notes in Computer Science (LNCS). Septembrie 2024 (Conferință ISI)

Data: 30.06.2025

Semnătură director,
Conf. Dr. Ing. Habil. Apostol Elena-Simona