6/5/2025

Polygraph

Detectarea “fake-news” în articole din presă

Militaru Ștefan

Cucu Mihnea

Oprescu Andrei

# 1.Ideea Proiectului

Scopul principal al proiectului polygraph este construirea unei aplicații care poate furniza date despre veridicitatea și legitimitatea unui articol ce îi pot fi utile unui utilizator. Deși recunoaștem că nu poate fi inventată o „soluție magică” care să spună cu certitudine garantată dacă un articol este complet adevărat, sperăm ca informațiile rezultate din analiza Polygrapgh să îi fie utile oricărui cititor de știri online.

Backlogul taskurilor care ne vor ajuta să ducem proiectul până la capăt este următorul:

1. Analiza textului știrii

• Extrage textul principal al articolului.

• Împarte conținutul în propoziții și analizează structura acestora.

2. Clasificare Fake News vs. Real News

• Algoritmi de machine learning care determină dacă articolul este fake sau real.

• Modele bazate pe NLP care detectează pattern-uri comune în știrile false (ex: exagerări, lipsa surselor, limbaj emoțional).

3. Verificarea surselor

• Compară site-ul articolului cu o bază de date de surse verificate (Media Bias/Fact Check, Open Sources).

• Avertizează utilizatorul dacă sursa este cunoscută pentru știri false.

4. Comparare cu știri de încredere

• Caută subiecte similare pe site-uri de știri verificate și compară faptele prezentate.

• Notifică utilizatorul dacă există diferențe semnificative între versiuni.

5. Analiză detaliată și raport explicativ

• Generează un raport cu argumente pro și contra pentru credibilitatea articolului.

• Explică ce factori au influențat scorul final.

• Evidențiază limbajul emoțional, lipsa surselor și alte trăsături problematice.

6. Interfață Web pentru utilizatori

• Un UI unde utilizatorii pot introduce un link sau un text de articol pentru analiză.

• Oferă un scor de credibilitate (ex: 0% – complet fake, 100% – foarte credibil).

7. Acces la sursele folosite pentru verificare

• Listează toate sursele folosite pentru verificare.

• Oferă link-uri către articolele relevante.

• Explică metodologia de selecție a surselor.

8. Sugestii de surse alternative

• Creează un mecanism care sugerează surse alternative de încredere.

• Prioritizează surse diverse pentru o perspectivă echilibrată.

9. Setări avansate pentru administratori

• Adaugă setări pentru pragurile scorului de credibilitate.

• Permite ajustarea sensibilității algoritmului.

10. Rezumat automatizat al articolelor

• Utilizează NLP pentru a extrage esențialul din articol.

• Afișează principalele puncte sub formă de listă.

11. Istoricul verificărilor

• Creează o pagină unde utilizatorul vede ultimele articole analizate.

• Permite ștergerea manuală a istoricului.

12. Raportarea analizelor incorecte

• Adaugă un buton „Raportează evaluarea” în UI.

• Creează un sistem de feedback pentru ajustarea modelului.

# Arhitectura aplicației

Funcționarea generală a aplicației se va realiza după următorii pași:

1. Utilizatorul introduce un link sau un text de articol.

2. Backend-ul preia articolul și îl analizează:

• Tokenizare și curățare text.

• Detectarea sursei și verificarea în baza de date.

• Compararea conținutului cu știri similare.

• Analiza sentimentului și clasificarea fake vs. real.

3. Returnează utilizatorului un scor de credibilitate + argumente.

Tehnologiile utilizate la realizarea aplicației vor fi:

* Python, pentru tokenizarea textului, analiza acestuia, rularea algoritmilor de ML
* Django, pentru interfața cu care va interacționa utilizatorul

# Aspecte Tehnice

În etapa de analiză a textului, am decis și să extragem câteva cuvinte cheie reprezentative care sunt menite să îl ajute pe utilizator să perceapă tonul textului și subiectele discutate de acesta. Pentru a realiza acest lucru ne vom folosi de frecvența cu care apar acești termeni și îi vom selecta pe aceia care apar disproporționat de des. Frecvența aparițiilor termenilor în articolul curent este comparată cu frecvența apariției lor în alte texte din engleză, cu ajutorul datasetului unigram\_freq, creditat în bibliografie.

#aici explicați și voi cum funcționează părțile implementate de voi#

În ceea ce privește analiza sursei textului ne vom folosi de baza de date de la allsides, secțiunea de Media Bias. Această bază de date clasifică sute de site-uri web/ autori de articole pe axa politică stânga – dreapta, bazându-se pe inputul utilizatorilor și furnizând astfel și un nivel de încredere. Întrucât scopul mai restrâns al proiectului nu justifică accesul prin licență la versiunea up to date online a acestei baze de date, vom folosi o versiune offline stocată în fișierul csv allsides\_data.

# Ce rămâne de implementat

Până la această etapă am finalizat primele 6 cerințe din backlog, deci pentru a finaliza proiectul mai avem de implementat a doua jumătate a acestuia. Va trebui să îmbunătățim funcționalitatea interfeței pentru a permite utilizatorilor să dea feedback, să afle mai multe despre analiza unui anumit articol sau să acceseze alte articole de încredere pe același subiect.

De asemenea va trebui să implementăm și câteva acțiuni pentru administratori care vor putea ajusta parametrii algoritmilor, în funcție de feedbackul dat de utilizatori.

# Bibliografie

* Fișierul cu frecvența apariției cuvintelor în engleză a fost preluat de pe Kaggle, de la Rachel Tatman, link: <https://www.kaggle.com/datasets/rtatman/english-word-frequency?resource=download>
* Versiunea offline a bazei de date allsides este preluată de pe github, <https://github.com/favstats/AllSideR/tree/master/data>
* Versiunea online a bazei de date, <https://www.allsides.com/media-bias>