

Sistem za prepoznavanje lažnih vijesti

Članovi tima

- Bojan Živanić SV61/2021

1. Opis problema

Motivacija

U digitalnom društvu veliki dio informacija dolazi sa interneta i društvenih mreža. Brzo širenje neprovjerenih i lažnih vijesti utiče na javno mnjenje, obrazovanje i donošenje odluka građana. Cilj projekta je razvoj ekspertnog sistema koji omogućava korisnicima da identifikuju potencijalno lažne vijesti koristeći pravila, transparentno rezonovanje i analizu obrazaca u sadržaju i izvorima.

Pregled problema

Većina postojećih rješenja koristi metode mašinskog učenja i NLP, što zahtijeva velike datasetove i složene modele. Takvi sistemi često nisu transparentni.

Predloženi sistem je rule-based, što omogućava:

- Vidljivost i objašnjenje odluka sistema
- Korišćenje forward i backward chaininga
- Analizu ponavljanja vijesti kroz CEP
- Standardizovan unos podataka putem templejta

Prednost sistema je jednostavna implementacija, transparentno rezonovanje i lako prilagođavanje novim pravilima i izvorima.

2. Metodologija rada

Ulazi (Input)

- Izvor vijesti (poznat, nepouzdan, nepoznat)
- Datum objave
- Naslov i sadržaj (ključne riječi, stil)
- Broj ponavljanja iste vijesti u kratkom vremenskom periodu
- Kontekst (drugi portali gde je vijest objavljena)

Izlazi (Output)

- Kategorija pouzdanosti:
 - Pouzdana vijest
 - Potencijalno lažna vijest
 - Sumnjiva vijest – potrebno dodatno proveriti
- Obrazloženje odluke sistema

Baza znanja

- Pravila o reputaciji izvora: poznat, nepouzdan, nepoznat
- Pravila o sadržaju: senzacionalističke fraze, pretjerani uzvičnici, netačne reference
- Pravila o ponavljanju vijesti (CEP)
- Template za uniforman unos vijesti, uključujući izvor, datum, stil i sadržaj

3. Forward chaining

Sistem koristi pravila koja se ulančavaju u tri nivoa:

Nivo 1 – Reputacija izvora

- Ako je izvor u listi poznatih portala, reputacija = *visoka*
- Ako je izvor u listi nepouzdatih portala ili je nepoznat, reputacija = *niska*

Nivo 2 – Analiza sadržaja

- Ako je reputacija *visoka* i nema senzacionalizma, preliminarna ocjena = *pouzdana vijest*
- Ako je reputacija *visoka* i ima senzacionalizam, preliminarna ocjene = *potencijalno lažna vijest*
- Ako je reputacija *niska* i nema senzacionalizma, preliminarna ocjena = *potencijalno lažna vijest*
- Ako je reputacija *niska* i ima senzacionalizam, preliminarna ocjena = *sumnjiva vijest*

Nivo 3 – CEP / Ponavljanja

1. Ako je preliminarna ocjena *pouzdana vijest* i *normalna_distribucija(vijest)*, finalna ocjena = *pouzdana vijest*
2. Ako je preliminarna ocjena *pouzdana vijest* i *koordinisano_sirenje(vijest)* ili *eksplozivno_sirenje(vijest)*, finalna ocjena = *sumnjiva vijest*
3. Ako je preliminarna ocjena *potencijalno lažna vijest* i *normalna_distribucija(vijest)*, finalna ocjena = *potencijalno lažna vijest*
4. Ako je preliminarna ocjena *potencijalno lažna vijest* i *koordinisano_sirenje(vijest)* ili *eksplozivno_sirenje(vijest)*, finalna ocjena = *sumnjiva vijest*
5. Ako je preliminarna ocjena *sumnjiva vijest*, finalna ocjena = *sumnjiva vijest* (bez obzira na CEP obrasce)

4. Backward chaining

Backward chaining mehanizam koristi hijerarhiju rekurzivnih upita koji provjeravaju postojanje već formiranih činjenica u radnoj memoriji. Sistem počinje od krajnjeg cilja (hipoteze) i rekurzivno provjerava niže nivoe dok ne dođe do elementarnih činjenica

Hijerarhija backward chaining upita:

- Nivo 0 - Glavna hipoteza
 - *da li je vijest pouzdana(vijest)* - konačna odluka sistema o pouzdanosti vijesti
- Nivo 1 - Primarni kriterijumi
 - *da li je izvor pouzdan(izvor)* - evaluacija reputacije izvora vijesti
 - *da li je sadržaj vjerodostojan(naslov, sadržaj)*- analiza vjerodostojnosti tekstualnog sadržaja
 - *da li su obrasci širenja normalni(vijest_id)*- procjena normalnosti distribucije vijesti
- Nivo 2 - Detaljna analize

Za izvor:

- *da li je izvor na listi pouzdanih(izvor)*
- *da li je izvor na listi nepouzdanih(izvor)*
- *da li je izvor pouzdan kroz ponavljanja(vijest)* - rekurzivna provjera izvora kroz lanac ponavljanja

Za sadržaj:

- *da li je naslov sažbi senzacionalizam(naslov)*
- *da li sadržaj ima rizične karakteristike(sadržaj)*

Za distribucije:

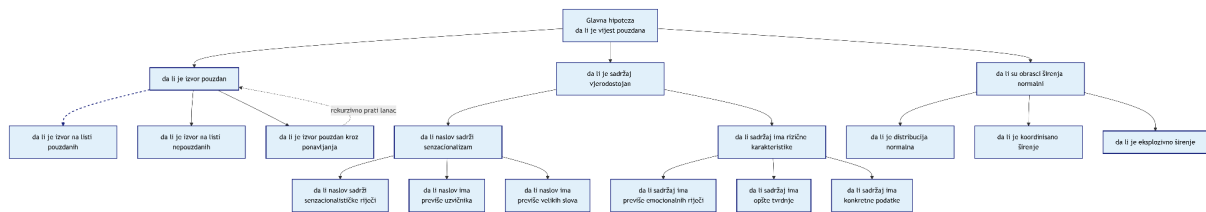
- *da li je događaj normalna distribucija(vijest_id)*
- *da li je događaj koordinisano širenje(vijest_id)*
- *da li je događaj eksplozivno širenje(vijest_id)*
- Nivo 3 - Granularna analiza

Analiza naslova:

- *da li naslov sadrži senzacionalističke riječi(naslov)*
- *da li naslov ima previše uzvičnika(naslov)*
- *da li naslov ima previše velikih slova(naslov)*

Analiza sadržaja:

- *da li sadržaj ima previše emocionalnih riječi(sadržaj)*
- *da li sadržaj ima opšte tvrdnje(sadržaj).*
- *da li sadržaj ima konkretne podatke(sadržaj)*



Query: *Da li je izvor pouzdan kroz ponavljanja(vijest)*

Rekurzivno prati lanac ponavljanja vijesti do originalnog izvora i provjerava postojanje činjenica o pouzdanosti svakog izvora u lancu. Algoritam:

1. Pronaći činjenicu *Ponavljanje(ponovljena vijest, originalna vijest)* za datu vijest
2. Ako postoji zapis o ponavljanju
 - a. Rekurzivno pozvati *da li je izvor pouzdan kroz ponavljanja(originalna vijest)*
 - b. Provjeriti *da li je izvor na listi pouzdanih(izvor originalne vijesti)*
3. Ako ne postoji zapis o ponavljanju
 - a. Provjeriti *da li je izvor na listi pouzdanih(izvor trenutne vijesti)*
4. Vratiti True samo ako postoje činjenice da su svi izvori u lancu pouzdani

5. CEP (Complex Event Processing)

CEP modul koristi se za analizu obrazaca pojavljivanja vijesti u realnom vremenu. CEP analizira kompleksne obrasce distribucije i širenja vijesti kroz detekciju specifičnih događaja. CEP modul koristi vremenske prozore i temporalne operatore za detekciju kompleksnih obrazaca širenja vijesti u realnom vremenu. Omogućava da se ocijeni da li širenje vijesti djeluje prirodno ili potencijalno manipulativno.

Pravila CEP-a:

Pravilo 1: Normalna distribucija

- Uslov: Vijest se pojavljuje na 2-3 pouzdana portala tokom dužeg vremenskog perioda(24-48h)
- Značenje: Prirodno širenje bez indikacija manipulacije
- Uticaj na ocjenu: Povećava pouzdanost vijesti

Pravilo 2: Koordinisano širenje

- Uslov: Vijest se pojavljuje na 3+ nepouzdana portala unutar 1 sat
- Značenje: Organizovano širenje sa potencijalnom manipulacijom
- Uticaj na ocjenu: Smanjuje pouzdanost vijesti

Pravilo 3: Eksplozivno širenje

- Uslov: Broj pojavljivanja se udvostručuje unutar 2 sata
- Značenje: Viralno širenje karakteristično za dezinformacije
- Uticaj na ocjenu: Smanjuje pouzdanost vijesti

Pravilo 4: Detekcija ponavljanja vijesti

- Uslov: Nova vijest(Vijest B) ima visoku sličnost(>85%) sa ranije objavljenom vijesti (Vijest A) u vremenskom periodu od 48h
- Akcija: Kreiranje činjenice *Ponavljanje*(*ponovljena vijest=B, originalna vijest=A*)
- Značenje: Vijest B predstavlja ponavljanje ranije objavljene vijesti A
- Uticaj: Omogućava backward chaining mehanizmu da prati lanac do originalnog izvora

6. Primjer konkretnog rezonovanja

Unos vijesti:

- Izvor: Vijesti365 (nepoznat portal)
- Naslov: „ŠOKANTNO! Ljekari sakrili istinu o vodi koju pijemo!“
- Datum: 9.9.2025. 14:30
- Sadržaj: „Ekskluzivno otkriće! Doktori skrivaju da obična voda iz slavine može izazvati nepopravljive zdravstvene probleme. Naučnici potvrđuju!“
- Kontekst: Detektovano 5 ponavljanja iste vijesti na različitim portalima

Koraci rezonovanja:

1. **Forward chaining - Nivo 1:** Sistem analizira izvor vijesti i utvrđuje nisku reputaciju (nepoznat izvor)
2. **Forward chaining - Nivo 2:** Analiza sadržaja otkriva senzacionalistički naslov
3. **Forward chaining - Nivo 3:** CEP detektuje *koordinisano širenje* (5 pojavljivanja/45 minuta)
4. **Backward chaining** pokušava dokazati hipotezu *vijest je pouzdana*, ali otkriva:
 - *da li je izvor pouzdan?* Rezultat: **False** (izvor nije pouzdan)
 - *da li je izvor pouzdan kroz ponavljanja?* Rezultat: CEP je detektovao *Ponavljanje*(*vijest365, originalna vijest*). Rekursivna provjera: originalna vijest je sa portala „BrzeVijesti“(nepouzdan izvor). Rezultat: **False**
 - *da li je sadržaj vjerodostojan?* Rezultat: **False** (naslov je senzacionalistički)
 - *da li su obrasci širenja normalni?* Rezultat: **False** (CEP je detektovao koordinisano širenje)
5. Konačna odluka: Vijest je sumnjiva

Izlaz sistema:

- Kategorija: *Sumnjiva vijest*
- Obrazloženje:
 - Izvor vijesti je nepoznat i nema utvrđenu reputaciju
 - Naslov sadrži elemente senzacionalizma
 - Detektovano koordinisano širenje
 - Lanac ponavljanja vodi do izvora sa niskom reputacijom