

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

Zasniva se na postavljanju ciljla i dokazivanju uslova unazad. Sistem koristi hijerarhijsku strukturu upita u kojoj svaka hipoteza ili podcilj vodi ka novim pitanjima dok se ne dođe do elementarnih činjenica u bazi znanja.

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)*- provjera liste pouzdanih izvora
- *da li je izvor na listi nepouzdanih(izvor)*-provjera liste nepouzdanih izvora

Za sadržaj:

- *da li je naslov sarži senzacionalizam(naslov)*- analiza stilskih karakteristika naslova
- *da li sadržaj ima rizične karakteristike(sadržaj)*-npr. Milioni ljudi su u opasnosti

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)*- npr. Šokantno
- *da li naslov ima previše uzvičnika(naslov)* - identifikacija prenaplašavanja
- *da li naslov ima previše velikih slova(naslov)* - detekcija CAPS LOCK efekta (preko 40% velikih slova)

Analiza sadržaja:

- *da li sadržaj ima previše emocionalnih riječi(sadržaj)*
- *da li sadržaj ima opšte tvrdnje(sadržaj)*- npr. Svi kažu ...
- *da li sadržaj ima konkretne podatke(sadržaj)*-npr. Stručnjaci kažu da je situacija zabrinjavajuća

Detaljnija struktura backward chaining mehanizma sa hijerarhijom funkcija i njihovim međusobnim povezivanjem prikazana je na sljedećim slikama koda:

```
1 # Nivo 0: Glavna hipoteza
2 ~ def da_li_je_vijest_pouzdana(vijest):
3     return (
4         da_li_je_izvor_pouzdan(vijest.izvor) and
5         da_li_je_sadrzaj_vjerodostojan(vijest.naslov, vijest.sadrzaj) and
6         da_li_su_obrasci_sirenja_normalni(vijest.id)
7     )
8
9 # Nivo 1: Glavni kriterijumi
10 ~ def da_li_je_izvor_pouzdan(izvor):
11     return (
12         da_li_je_izvor_na_listi_pouzdanih(izvor) and
13         not da_li_je_izvor_na_listi_nepouzdanih(izvor)
14     )
15
16 ~ def da_li_je_sadrzaj_vjerodostojan(naslov, sadrzaj):
17     return (
18         not da_li_je_naslov_senzacionalisticki(naslov) and
19         not da_li_sadrzaj_ima_rizicne_karakteristike(sadrzaj)
20     )
21
22 ~ def da_li_su_obrasci_sirenja_normalni(vijest_id):
23     return (
24         da_li_je_dogadjaj_normalna_distribucija(vijest_id) and
25         not da_li_je_dogadjaj_koordinisano_sirenje(vijest_id) and
26         not da_li_je_dogadjaj_eksplozivno_sirenje(vijest_id)
27     )
28
29 # Nivo 2: Analiza izvora
30 ~ def da_li_je_izvor_na_listi_pouzdanih(izvor):
31     return izvor in POUZDANI_IZVORI
32
33 ~ def da_li_je_izvor_na_listi_nepouzdanih(izvor):
34     return izvor in NEPOUZDANI_IZVORI
35
36 # Nivo 2: Analiza sadržaja - NASLOV
37 ~ def da_li_je_naslov_senzacionalisticki(naslov):
38     return (
39         da_li_naslov_sadrzi_senzacionalisticke_rijeci(naslov) or
40         da_li_naslov_ima_previše_uzvicih(naslov) or
41         da_li_naslov_ima_previše_velikih_slova(naslov)
42     )
43
44 # Nivo 3: Detaljna analiza naslova
45 ~ def da_li_naslov_sadrzi_senzacionalisticke_rijeci(naslov):
46     return any(rijec in naslov.lower() for rijec in SENZACIONALISTICKE_RUJECI)
47
48 ~ def da_li_naslov_ima_previše_uzvicih(naslov):
49     return naslov.count('!') >= 2
50
51 ~ def da_li_naslov_ima_previše_velikih_slova(naslov):
52     return sum(1 for c in naslov if c.isupper()) > len(naslov) * 0.4
53
54 # Nivo 2: Analiza sadržaja - SADRŽAJ
55 ~ def da_li_sadrzaj_ima_rizicne_karakteristike(sadrzaj):
56     return (
57         da_li_sadrzaj_ima_previše_emocionalnih_rijeci(sadrzaj) or
58         da_li_sadrzaj_ima_opšte_tvrđnje(sadrzaj) or
59         da_li_sadrzaj_nema_konkretnih_podatke(sadrzaj)
60     )
61
62 # Nivo 3: Detaljna analiza sadržaja
63 ~ def da_li_sadrzaj_ima_previše_emocionalnih_rijeci(sadrzaj):
64     emocionalne_rijeci = ["strah", "panika", "kriza", "katastrofa", "skandal", "šok", "užas", "strašno"]
65     sadrzaj_lower = sadrzaj.lower()
66     return sum(1 for rijec in emocionalne_rijeci if rijec in sadrzaj_lower) >= 3
67
68 ~ def da_li_sadrzaj_ima_opšte_tvrđnje(sadrzaj):
69     opce_tvrđnje = ["svi znaju", "svi kažu", "opće je poznato", "očigledno je", "svi misle", "svi vide"]
70     sadrzaj_lower = sadrzaj.lower()
71     return any(tvrđnja in sadrzaj_lower for tvrđnja in opce_tvrđnje)
72
73 ~ def da_li_sadrzaj_nema_konkretnih_podatke(sadrzaj):
74     sadrzaj_lower = sadrzaj.lower()
75     return (
76         not any(rijec in sadrzaj_lower for rijec in ["prema", "izvor", "rekao", "izjavio", "navodi"]) and
77         not any(c.isdigit() for c in sadrzaj)
78     )
79
80 # Nivo 2: CEP analiza
81 ~ def da_li_je_dogadjaj_normalna_distribucija(vijest_id):
82     return cep_da_li_je_normalna_distribucija(vijest_id)
83
84 ~ def da_li_je_dogadjaj_koordinisano_sirenje(vijest_id):
85     return cep_da_li_je_koordinisano_sirenje(vijest_id)
86
87 ~ def da_li_je_dogadjaj_eksplozivno_sirenje(vijest_id):
88     return cep_da_li_je_eksplozivno_sirenje(vijest_id)
```

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

6. Primjer konkretnog rezonovanja

Unos vijesti:

- Izvor: Nepoznat portal
- Naslov: „ŠOKANTNO! Ljekari sakrili istinu o vodi koju pijemo!“
- Datum: Juče
- Ponavljanja: 5 puta (na nepouzdanim portalima u roku od 45 minuta)

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 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: Vijest dolazi sa nepoznatog izvora, ima senzacionalistički naslov i detektovano je koordinisano širenje na nepouzdanim portalima u kratkom vremenskom periodu.