

În contextul în care rețelele sociale au devenit mediul principal de exprimare și propagare a informației, capacitatea de a analiza și extrage date utile din conținutul postat în cadrul acestor rețele capătă din ce în ce mai multă importanță. Aceste informații stau la baza unei plaje largi de aplicații practice într-o varietate mare de domenii cum ar fi marketing[1], sisteme de recomandare[2], predicția burselor[3] sau politică[4][5]. Prezenta lucrare prezintă o serie de metode dezvoltate pentru a analiza opiniile utilizatorilor unei rețele sociale pe baza comportamentului afișat în cadrul acestora. Rețeaua socială avută în vedere este Twitter, dar sistemele prezentate pot fi adaptate și pentru alte rețele sociale.

Obiectivele propuse sunt următoarele:

1. Dezvoltarea unui sistem care, având o serie de informații despre comportamentul unui utilizator, poate prezice cu o anumită precizie opinia acestuia cu privire la un subiect.
2. Dezvoltarea unui sistem care, având la dispoziție un text, poate estima opinia creatorului cu privire la un anumit subiect. În text nu este necesar să se facă o referire imediată la subiectul vizat. De exemplu dacă subiectul este Hillary Clinton, iar textul este "I like Donald Trump #makeAmericaGreatAgain", un observator uman poate să concluzioneze că utilizatorul nu este un simpatizant al lui Hillary Clinton datorită faptului că Donald Trump și Hillary Clinton sunt adversari politici. Obiectivul este de a face astfel de determinări automat.

Sistemele dezvoltate se bazează pe învățarea în prealabil a unor reprezentări distribuite pentru fiecare utilizator. Practic, fiecare utilizator are asociat un vector de dimensiune  $N$ , iar spațiul  $N$ -dimensional obținut are proprietatea că utilizatorii cu opinii similare au asociați vectori apropiați cu privire la o anumită metrică. Acești vectori vor reprezenta de fapt o codificare numerică a opiniilor fiecărui utilizator. Utilizatorii astfel codificați pot fi folosiți ca intrări în cadrul unor sisteme de clasificare care pot estima anumite aspecte legate de aceștia.

Sistemul propus pentru a determina opiniile utilizatorilor cu privire la un anumit subiect generează un număr de utilizatori a căror opinie este cunoscută și construiește o mașină cu vectori suport care folosește vectorii asociați acestor utilizatori pentru a genera un model cu ajutorul căruia să fie clasificați ceilalți utilizatori.

Pentru a determina opinia exprimată într-un anumit text este folosită o rețea neuronală. Această rețea primește ca intrare un text și scopul este de a estima opinia creatorului aceluși text referitor la subiectul avută în vedere. Particularitatea sistemului

dezvoltat este faptul că datele de antrenament sunt obținute automat, pe baza clasificărilor utilizatorilor obținute anterior, fără a fi nevoie de un set de date etichetat manual, ceea ce permite crearea unui set de antrenament foarte mare.

Precizia rezultatelor obținute este măsurată pe baza unor date de test. În cadrul testelor efectuate, precizia obținută pentru prima problema ajunge pentru anumite subiecte la 99%, iar pentru cea de-a doua problema propusă ajunge la 86%. De asemenea au fost atașate o serie de ilustrații care facilitează vizualizarea rezultatelor.