**Analiza sentimentelor exprimate în recenziile telefoanelor Huawei de pe Amazon**

**Studenți: Măgdălin Claudiu-Ștefan**

**Mewis Iohan Ștefan**

**Mihăilă Andrei**

**Introducere**

**Sentiment Analysis**sau **Opinion Mining** este o ramură a domeniului de Natural Language Processing (NLP) ce se ocupă cu studiul opiniilor, sentimentelor, evaluărilor, atitudinilor, emoțiilor și caracteristicele acestora, direcționate spre anumite entități precum produse, organizații, indivizi, evenimente, etc. Nu a existat un interes deosebit pentru această disciplină înainte de anul 2000, dar odată cu răspândirea aplicațiilor comerciale (onlinesauoffline*),* perspectiva asupra analizei opiniilor s-a schimbat în mod radical. E și pentru prima dată în istorie când există o bază de date dogmatică consistentă, alcătuită cu ajutorul rețelelor de socializare. Nu e atât de surprinzătoare schimbarea de perspectivă ținând cont de aplicațiile sale în domenii politice, sociologice, economice și de popularitatea de care se bucură socializarea online în zilele noastre.

**Descrierea proiectului**

Pe măsură ce accesul la informații devine mai ușor, tot mai mulți consumatori vor căuta informații despre produse de la alți consumatori, în afară de informațiile furnizate de vânzător. Recenziile și evaluările transmise de consumatori sunt exemple de astfel de informații și au devenit deja o parte integrantă a procesului de decizie de cumpărare a clienților.

În proiectul pe care îl vom realiza vom analiza o serie de recenzii postate din SUA pe Amazon pentru telefoanele Huawei.

**Motivația**

Conform unui studiu realizat în anul 2018 (Local Consumer Review Survey), 86% dintre (posibilii) clienți citesc review-uri înainte de a lua decizia de cumpărare, iar 91% dintre aceștia au încredere în ele la fel de mult ca într-o recomandare venită din partea unui prieten. Considerăm importantă tema abordată deoarece tot mai multe companii se axează pe procese automate de statistică, pe baza recenziilor obținute din partea clienților, în încercarea de a oferi acestora un produs cât mai bun. Pe baza considerentelor menționate, avantajele pe care tema realizată de noi le aduce sunt:

* Îmbunatatirea produselor obținute;
* Acumularea de informații despre publicul țintă al serviciilor oferite;
* Raportarea la competiție.

**Literatura de specialitate**

Analiza sentimentelor referă la utilizarea prelucrării limbajului natural, analizei de text precum și lingvisticii computaționale pentru a identifica și extrage informații cu caracter subiectiv din materiale sursă. La modul general analiza sentimentelor urmărește să identifice atitudinea unei anumite persoane referitoare la o anumită topică sau la contextul unui document. Atitudinea poate fi propria judecată, starea emoțională sau comunicarea emoțională intenționată de autor.

Primele cercetări în acest domeniu îi aduc în prim-plan pe Turney[[1]](#footnote-1) şi Pang[[2]](#footnote-2), care au aplicat diferite metode pentru a determina polaritatea recenziilor unor produse, respectiv filme. Activitatea s-a desfăşurat la nivel de document. Polaritatea unui document se poate clasifica de asemenea pe o scară multiplă, direcţie adoptată de către Pang şi Snyder[[3]](#footnote-3) (printre alţii). Pang a extins sarcina de bază privind clasificarea unei recenzii de film de la a fi pozitivă sau negativă la estimarea unor evaluări pe o scară de la 3 la 4 stele, în timp ce Snyder a desfăşurat o analiză în profunzime asupra recenziilor unor restaurante, estimând evaluări privind diferite aspecte ale restaurantului dat, precum calitatea mâncării şi atmosfera (pe o scară de la una la cinci stele).

O metodă diferită de determinare a sentimentului este folosirea unui sistem de clasificare în care pentru cuvintele asociate frecvent cu un sentiment negativ, neutru sau pozitiv, se stabileşte un număr pe o scară de la -10 la 10 (cel mai negativ către cel mai pozitiv) şi atunci când un fragment de text nestructurat este analizat utilizând NLP (Natural Language Processing), conceptele rezultate în urma analizei vor fi la rândul lor analizate pentru înţelegerea cuvintelor şi modului în care acestea se raportează la conceptele respective

**Provocări**

Una dintre cele mai dificele provocări de cercetare în NLP este identificarea declarațiilor sarcastice. O altă provocare în analiza sentimentului este prezența cuvintelor de negare care inversează întotdeanu orientarea sentimentului.

Analiza sentimentelor utilizând recenziile telefoanelor prezintă provocări care nu sunt de obicei nu sunt găsite în timpul analizei sentimentelor de text formal. Recenziile conțin adesea joargoane și cuvinte scrise rapid și prescurtat. De asemenea, recenziile sunt texte informale, prin urmare există multe cuvinte ce pot implica o orientare diferită în funcție de grupa de vârstă a autorului.

**Arhitectura (schema)**

Clasificarea polaritatii sentimentelor

Curatarea bazei de date

Identificarea sentimentelor

Descrierea statistica a variabilelor numerice

1. Turney, 2002] Turney, Peter. "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews". Proceedings of the Association for Computational Linguistics. pp. 417–424 [↑](#footnote-ref-1)
2. 0 [Pang, 2002] Pang, Bo; Lee, Lillian; Vaithyanathan, Shivakumar. "Thumbs up? Sentiment Classification using Machine Learning Techniques". Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 79–86. [↑](#footnote-ref-2)
3. [Snyder, 2007] Snyder, Benjamin; Barzilay, Regina. "Multiple Aspect Ranking using the Good Grief Algorithm". Proceedings of the Joint Human Language Technology/North American Chapter of the ACL Conference (HLT-NAACL). pp. 300–307. [↑](#footnote-ref-3)