**Universitatea Tehnică „Gheorghe Asachi” din IAŞI**

**Facultatea de Automatică şi Calculatoare**

**Domeniul:** ***Calculatoare şi Tehnologia Informaţiei***

**Specializarea:** ***Calculatoare/Tehnologia Informaţiei***

**Platformă web pentru analiza și vizualizarea emoțiilor pe site-uri de socializare**

Lucrare de Licenţă

Absolvent

Alexandru Ivanov

Coordonator ştiinţific

Ş.l. dr. Iulian PETRILA

# Iaşi, 2024

# DECLARAŢIE DE ASUMARE A AUTENTICITĂŢII

# LUCRĂRII DE LICENŢĂ

Subsemnatul(a) Prenume NUME,

legitimat(ă) cu BI/CI..

autorul lucrării **Titlu**

elaborată în vederea susţinerii examenului de finalizare a studiilor de licenţă  
organizat de către Facultatea de Automatică şi Calculatoare din cadrul Universităţii Tehnice „Gheorghe Asachi” din Iaşi, sesiunea iulie a anului universitar 2016-2017, luând în considerare conţinutul Art. 34 din Codul de etică universitară al Universităţii Tehnice „Gheorghe Asachi” din Iaşi (Manualul Procedurilor, UTI.POM.02 – Funcţionarea Comisiei de etică universitară), declar pe proprie răspundere, că această lucrare este rezultatul propriei activităţi intelectuale, nu conţine porţiuni plagiate, iar sursele bibliografice au fost folosite cu respectarea legislaţiei române (legea 8/1996) şi a convenţiilor internaţionale privind drepturile de autor.

Data Semnătura

Zi – Lună – An

# Rezumat

(1-2 pagini)

# Cuprins

[**Introducere** 1](#_Toc475694174)

[**Fundamentarea teoretică şi tehnologiile utilizate** 2](#_Toc475694175)

[**Componentele şi caracteristicile metodei sau/şi a aplicaţiei propuse** 3](#_Toc475694176)

[**Implementarea, gestiunea, testarea şi utilizarea metodei/aplicaţiei** 4](#_Toc475694177)

[**Concluzii** 5](#_Toc475694178)

[**Bibliografie** 6](#_Toc475694179)

[**Anexe** 7](#_Toc475694180)

[***Anexa1 – Detalii elemente teoretice*** 7](#_Toc475694181)

[***Anexa2 – Codul sursă*** 7](#_Toc475694182)

# Introducere

În era digitală contemporană, rețelele de socializare au devenit parte integrantă a vieților noastre, oferind o platformă pentru comunicare, partajare de informații și exprimarea emoțiilor. Aceste platforme generează un volum vast de date textuale care pot fi analizate pentru a înțelege mai bine emoțiile utilizatorilor. Analiza sentimentelor din aceste date a evoluat semnificativ de-a lungul timpului, trecând prin diferite etape și abordări, fiecare cu propriile sale avantaje și limitări.

La începutul dezvoltării analizelor sentimentelor, abordările lexicon-based au fost cele mai utilizate. Aceste metode implică utilizarea unor dicționare de sentimente, care asociază cuvinte individuale cu scoruri de polaritate (pozitiv, negativ, neutru). Primele implementări notabile includ AFINN și SentiWordNet, care au furnizat baze de date extinse de cuvinte asociate cu scoruri de sentiment [1]. Avantajul principal al acestor metode este simplitatea și interpretabilitatea lor, dar ele au limitări semnificative în gestionarea complexității și a nuanțelor emoționale din textele mai lungi și mai elaborate.Un avans semnificativ în această direcție a fost realizat prin dezvoltarea VADER (Valence Aware Dictionary and sEntiment Reasoner). VADER, integrat în biblioteca NLTK (Natural Language Toolkit), este conceput special pentru analiza sentimentelor din datele generate pe rețelele sociale. Acesta utilizează un dicționar de sentimente, completat de reguli heuristice care capturează intensitatea și nuanțele contextuale ale sentimentelor. Studiile au demonstrat că VADER poate atinge performanțe comparabile cu modelele bazate pe învățare automată în anumite contexte, având avantajul de a fi rapid și ușor de implementat [2].Cu evoluția tehnologiei, metodele de învățare automată au devenit din ce în ce mai populare în analiza sentimentelor. Algoritmi precum SVM (Support Vector Machines) și Naive Bayes au fost printre primele tehnici utilizate pentru a clasifica sentimentele pe baza trăsăturilor extrase din text [3]. Aceste metode au demonstrat o acuratețe superioară comparativ cu abordările lexicon-based, deoarece pot învăța din datele etichetate și pot captura relațiile complexe dintre cuvinte.

Rețelele neuronale au adus o revoluție în domeniu datorită capacității lor de a învăța reprezentări de caracteristici complexe și nuanțate. Rețelele neuronale recurente (RNN) și variantele lor, cum ar fi LSTM (Long Short-Term Memory) și GRU (Gated Recurrent Units), au permis capturarea dependențelor pe termen lung în texte, oferind o înțelegere mai profundă a contextului .Un salt major în analiza sentimentelor a fost realizat prin dezvoltarea modelelor pre-antrenate, cum ar fi BERT (Bidirectional Encoder Representations from Transformers) și GPT-3. Aceste modele utilizează arhitecturi Transformer și sunt pre-antrenate pe corpuri masive de date textuale, permițându-le să înțeleagă contextul și nuanțele subtile ale limbajului [4]. BERT, de exemplu, este antrenat bidirecțional, capturând contextul din ambele părți ale unui cuvânt într-o propoziție, ceea ce îi permite să înțeleagă mai bine sentimentele complexe și contextuale. GPT-3, cu capacitatea sa de a genera text coerent și contextual relevant, poate fi utilizat pentru a analiza și genera texte cu diverse tonuri emoționale. Aceste modele oferă o acuratețe superioară și pot fi adaptate pentru diverse domenii și aplicații.

Analiza sentimentelor pe rețelele sociale a devenit un subiect de interes major datorită cantității mari de date generate de utilizatori și a relevanței acestor date pentru diverse domenii. Una dintre cele mai comune aplicații ale analizei sentimentelor pe rețelele sociale este monitorizarea opiniei publice. Studiile au demonstrat că analiza sentimentelor poate oferi perspective valoroase asupra opiniei publice, tendințelor sociale, prezice rezultatele electorale și chiar predictibilității pieței financiare [5]. Companiile utilizează analiza sentimentelor pentru a monitoriza feedback-ul clienților pe platformele de socializare și pentru a îmbunătăți serviciile și produsele oferite. Evaluarea recenziilor și comentariilor postate de utilizatori poate oferi informații valoroase despre punctele forte și slabe ale unui produs sau serviciu, permițând companiilor să reacționeze rapid la problemele identificate. De asemenea, analiza sentimentelor poate fi utilizată pentru a identifica și analiza tendințele sociale emergente. De exemplu, cercetătorii pot analiza conversațiile online pentru a detecta schimbările în atitudinile față de anumite subiecte, cum ar fi sănătatea publică, schimbările climatice sau mișcările sociale. Aceste informații pot fi folosite pentru a dezvolta politici și strategii mai bine informate.

Un exemplu notabil este studiul realizat de Thelwall, care a explorat sentimentele exprimate pe platforma MySpace folosind o combinație de metode lexicon-based și statistici textuale [6]. De asemenea, Pang și Lee au fost pionieri în utilizarea recenziilor de filme pentru a dezvolta și testa modele de analiză a sentimentelor, demonstrând eficiența tehnicilor de învățare automată în acest domeniu [7]. Acestea au studiat doar intensitățile emoțiilor într-un sens liniar, de la negativ la pozitiv si stare neutră. Pornind de la acest fapt, lucrarea își propune să facă o distincție mai largă a emoțiilor exprimate în cadrul ideilor/discuțiilor captate din rețele de socializare consacrate. Abordările tradiționale în analiza sentimentelor au studiat intensitățile emoțiilor într-un sens liniar, variind de la negativ la pozitiv și incluzând o stare neutră. Deși aceste metode au furnizat o bază solidă pentru înțelegerea emoțiilor în mediul online, ele sunt limitate în captarea complexității și diversității emoțiilor umane. În acest context, lucrarea de față își propune să extindă această abordare prin realizarea unei distincții mai largi și mai nuanțate a emoțiilor exprimate în discuțiile și ideile captate din rețelele de socializare consacrate. Prin utilizarea tehnicilor de procesare a limbajului natural (NLP), această cercetare vizează identificarea și clasificarea unui spectru mai larg de emoții, oferind de asemenea și o vizualizare amplă pentru identificarea de pattern-uri și extragerea de informații importante despre domeniile aferente. Astfel, vizualizarea rezultatelor prin grafice și diagrame interactive va oferi utilizatorilor o imagine clară și intuitivă a datelor analizate, evidențiind modele recurente și anomalii.

# Capitolul 1.Fundamentarea teoretică şi tehnologiile utilizate

Analiza emoțiilor pe site-urile de socializare reprezintă un domeniu de cercetare emergent și deosebit de relevant în contextul digital actual. Cu milioane de utilizatori activi zilnic pe platforme precum Facebook, Twitter, Instagram și altele, volumul de date generate este enorm. Aceste date includ opinii, gânduri, reacții și emoții ale utilizatorilor, oferind o sursă bogată de informații pentru diverse scopuri. Pentru a asigura o înțelegere clară a subiectului, este important să definim conceptele cheie utilizate în analiza emoțiilor pe site-urile de socializare.

Emoțiile reprezintă reacții psihologice și fiziologice complexe la stimuli interni sau externi. Ele sunt esențiale pentru supraviețuire și adaptare, influențând comportamentul și deciziile noastre zilnice. Emoțiile pot fi clasificate în diferite moduri. Conform teoriei lui Paul Ekman [8], există șase emoții de bază universale recunoscute prin expresii faciale specifice: fericire, tristețe, frică, furie, surpriză și dezgust. Aceste emoții sunt considerate universale deoarece sunt recunoscute în toate culturile umane. Modelul circumplex al emoțiilor, propus de James Russell [9], sugerează că emoțiile pot fi reprezentate într-un spațiu bidimensional, definit de două axe: valența (pozitiv-negativ) și excitația (ridicat-scăzut). Acest model ajută la înțelegerea modului în care emoțiile se pot combina și varia.

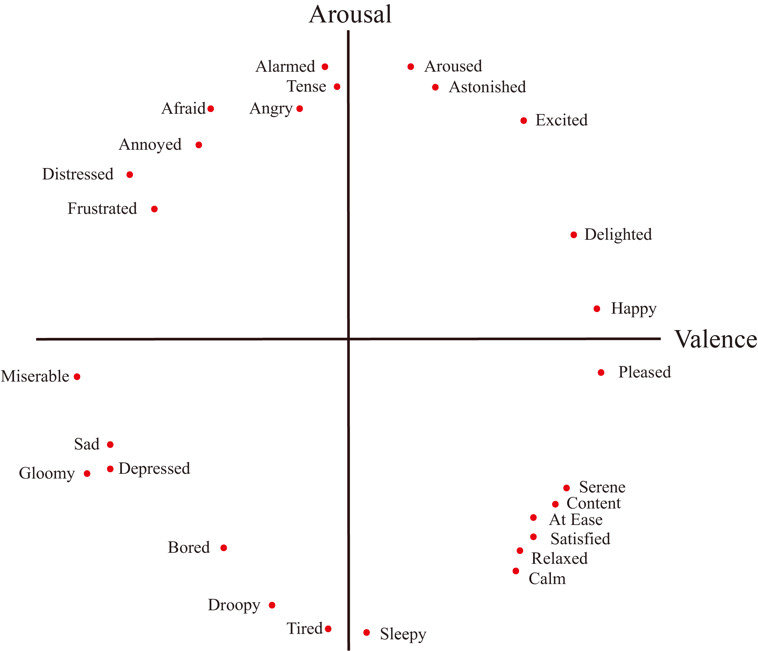


Figura 1.1. Modelul circumplex al lui Russel.

Impactul social media asupra societății este vast, influențând modul în care consumăm știri, formăm opinii, interacționăm cu branduri și participăm la discuții sociale și politice. Social media reprezintă o categorie vastă de platforme online care permit utilizatorilor să creeze, să partajeze și să interacționeze cu conținut generat de alți utilizatori. Aceste platforme au revoluționat modul în care comunicăm, interacționăm și consumăm informații. Există mai multe tipuri de platforme de social media, fiecare cu propriile caracteristici și scopuri. Platforme precum Facebook și LinkedIn care facilitează conectarea și interacțiunea între utilizatori prin crearea de profiluri personale, grupuri de interese și pagini de afaceri. Twitter, unde utilizatorii pot posta mesaje scurte (tweet-uri) și pot urmări actualizările altor utilizatori. Aceasta este cunoscută pentru rapiditatea și concizia comunicării. Un alt tip de platforme sunt cel de partajare a conținutului media. Instagram, YouTube și TikTok permit utilizatorilor să creeze, să distribuie și să vizualizeze conținut media, cum ar fi fotografii, videoclipuri și transmisiuni live. În plus, forumurile și comunitățile online contribuie semnificativ în discuțiile pe subiecte de interes și facilitează găsirea răspunsurilor pentru diferite probleme. Platforme precum Reddit și Quora sunt astfel structurate în sub-forumuri (subreddits) pentru discuții axate pe domenii singulare de discuție.

## Analiza sentimentelor

Analiza de sentiment este un domeniu care combină tehnici din procesarea limbajului natural (NLP), învățarea automată (ML) și mining-ul de text pentru a identifica și extrage informații subiective din texte. Scopul principal al analizei de sentiment este de a determina atitudinea autorului față de un anumit subiect, aceasta putând fi pozitivă, negativă sau neutră.

Metodele lexicon-based pentru analiza de sentiment se bazează pe utilizarea dicționarelor de cuvinte asociate cu sentimente specifice pentru a evalua textele. Aceste metode sunt directe și utile pentru a determina polaritatea unui text (pozitiv, negativ, neutru) fără a necesita date etichetate pentru antrenare. Principiul de bază al acestor metode este de a compara fiecare cuvânt din text cu un lexicon de sentimente și de a calcula un scor total de sentiment pe baza prezenței și frecvenței cuvintelor din lexicon. Construirea lexiconului de sentimente implică crearea unei liste de cuvinte și expresii asociate cu sentimente pozitive, negative sau neutre. Exemple de lexiconuri includ SentiWordNet, AFINN și VADER. SentiWordNet atribuie scoruri de polaritate fiecărui cuvânt din WordNet, AFINN oferă scoruri numerice care indică sentimentul fiecărui cuvânt, de la -5 (foarte negativ) la +5 (foarte pozitiv), iar VADER este conceput special pentru analiza de sentiment pe texte din social media, atribuind scoruri de intensitate sentimentelor. Funcționarea metodelor lexicon-based implică mai multe etape. Textul este mai întâi tokenizat, adică împărțit în unități mai mici (cuvinte sau fraze). Fiecare cuvânt din text este apoi comparat cu lexiconul de sentimente. Cuvintele găsite în lexicon sunt evaluate și li se atribuie un scor de sentiment. Scorurile tuturor cuvintelor din text sunt cumulate pentru a obține un scor total de sentiment. Cuvintele pozitive și negative sunt identificate, iar scorul total reflectă sentimentele exprimate în text. Cuvintele precum "nu" sau "niciodată" (negări) și "foarte" sau "extrem de" (amplificatori) sunt identificate pentru a ajusta scorurile de sentiment ale cuvintelor învecinate. Negările pot inversa sentimentul unui cuvânt, în timp ce amplificatorii cresc intensitatea sentimentului. Metodele lexicon-based au mai multe avantaje, printre care se numără simplitatea și ușurința în implementare, eficiența pentru texte scurte și bine structurate, și faptul că nu necesită date etichetate pentru antrenare. Totuși, aceste metode au și dezavantaje semnificative, inclusiv limitările în captarea nuanțelor complexe ale limbajului natural, sensibilitatea redusă la context și sarcini complexe, și dependența de calitatea și acoperirea lexiconului utilizat.

A diagram of a learning process

Description automatically generatedPe de altă parte, metodele de inteligență artificială pentru analiza de sentiment utilizează algoritmi de învățare automată și învățare profundă pentru a analiza textele și a determina sentimentele exprimate. Aceste metode implică antrenarea modelelor pe seturi mari de date etichetate pentru a învăța tipare și a prezice sentimentele în texte noi. Colectarea datelor și etichetarea implică adunarea de texte din diverse surse și etichetarea lor manuală pentru a indica sentimentele. Preprocesarea datelor este esențială pentru a elimina zgomotul și a standardiza formatul, incluzând tokenizarea, eliminarea stop word-urilor, stemming și lematizare. Caracteristicile relevante sunt extrase din texte, cum ar fi frecvența cuvintelor, n-gramme și reprezentările vectoriale ale cuvintelor (embeddings). Modelele de învățare automată, cum ar fi Naive Bayes și SVM, sau rețelele neuronale, cum ar fi LSTM și BERT [4], sunt antrenate pe seturile de date etichetate. Metodele de inteligență artificială oferă avantaje semnificative, inclusiv capacitatea de a captura nuanțele complexe ale limbajului natural și de a se adapta la diverse contexte. Aceste metode pot oferi o acuratețe mai mare și sunt capabile să învețe din seturi mari de date. Totuși, metodele de IA necesită date etichetate pentru antrenare, ceea ce poate fi costisitor și consumator de timp.

Figura 1.2. Metode de analiză a sentimentelor din text

### 1.1.1. VADER Lexicon

Printre diversele metode utilizate pentru această analiză, lexiconul VADER (Valence Aware Dictionary and sEntiment Reasoner) se evidențiază prin capacitatea sa de a analiza în mod precis sentimentele exprimate în textele de pe rețelele de socializare și alte forme de comunicare informală. VADER este un lexicon și un model de analiză de sentiment special conceput pentru a trata limbajul folosit în rețelele de socializare. Acesta a fost dezvoltat de C.J. Hutto și Eric Gilbert [2]. VADER se bazează pe o combinație de reguli și o listă de cuvinte cu scoruri de polaritate predefinite, care au fost validate empiric prin multiple studii. VADER a fost creat pentru a analiza eficient textele scurte, cum ar fi tweet-urile, comentariile de pe Facebook și recenziile online. Acesta ia în considerare atât polaritatea (pozitivă, negativă, neutră) cât și intensitatea sentimentelor exprimate. Lexiconul VADER conține aproximativ 7,500 de cuvinte, fiecare cu un scor de sentiment asociat care variază de la -4 (foarte negativ) la +4 (foarte pozitiv). Aceste scoruri au fost determinate printr-un proces de crowdsourcing, unde oameni reali au evaluat sentimentele asociate fiecărui cuvânt.

Funcționarea VADER se bazează pe câteva principii esențiale. Fiecare cuvânt din lexiconul VADER are un scor de sentiment predefinit. Atunci când VADER analizează un text, acesta compară fiecare cuvânt din text cu lexiconul pentru a atribui un scor de sentiment. Acesta utilizează reguli heuristice pentru a ajusta scorurile de sentiment pe baza contextului. De exemplu, amplificatorii precum "foarte" sau "extrem de" cresc intensitatea sentimentului asociat unui cuvânt. Negările precum "nu" sau "niciodată" inversează polaritatea sentimentului. VADER ia în considerare utilizarea punctuației și a emoticoanelor. De exemplu, un text cu multe semne de exclamare poate avea un sentiment mai intens, iar emoticoanele sunt tratate ca indicatori direcți ai sentimentelor. Cuvintele scrise cu majuscule sunt considerate mai intense. De exemplu, "BINE" are un scor de sentiment mai mare decât "bine".

VADER prezintă mai multe avantaje care îl fac ideal pentru analiza de sentiment a textelor de pe rețelele de socializare și alte forme de comunicare informală. VADER a fost validat empiric și s-a dovedit a fi la fel de precis, dacă nu mai precis, decât metodele de învățare automată supravegheate pentru analiza de sentiment. Fiind un model bazat pe reguli, este simplu de implementat și rulează rapid, făcându-l ideal pentru analize în timp real. Acesta fost special conceput pentru a înțelege limbajul informal și colocvial utilizat pe rețelele de socializare, inclusiv utilizarea jargonului, a acronimelor și a emoticoanelor. Spre deosebire de modelele de învățare automată, VADER nu necesită un set de date etichetat pentru antrenare, fiind gata de utilizare imediat după instalare. Cu toate că are multe avantaje, există și câteva limitări de care utilizatorii trebuie să fie conștienți. Deoarece acest model se bazează pe un set predefinit de reguli și scoruri, poate să nu fie la fel de adaptabil ca modelele de învățare automată care pot învăța și evolua pe baza unor noi seturi de date. Deși gestionează bine limbajul informal, poate întâmpina dificultăți în tratarea unor construcții lingvistice foarte complexe sau ambigue. VADER analizează fiecare text în mod izolat și nu ia în considerare contextul mai larg în care textul a fost scris, ceea ce poate duce la interpretări greșite în anumite cazuri.

### 1.1.2 Metode bazate pe rețele neuronale

# Componentele şi caracteristicile metodei sau/şi a aplicaţiei propuse

Descrierea problemei, căilor de soluţionare, componentele, caracteristicile etc. (de preferinţă într-un capitol dar nu mai mult de două).

# Implementarea, gestiunea, testarea şi utilizarea metodei/aplicaţiei

Detalii cu privire la modul concret în care a fost implementată, testată, verificată aplicaţia sau metoda sau tehnologia propusă, despre cum ar trebui un utilizator sa o folosească, despre utilitatea, limitele şi perspectivele ei etc. (1 capitol sau maxim 2).

# Concluzii

Succint, de prezentat în cate o frază elementele esenţial distincte ale temei abordate (minim o pagina, maxim două).

# Bibliografie

1. *Esuli, Andrea, and Fabrizio Sebastiani***.** "SentiWordNet: A publicly available lexical resource for opinion mining." *Proceedings of LREC*. Vol. 6. 2006.
2. *Hutto, C., și Eric Gilbert***.** "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, no. 1, 2014, pp. 216-225, Disponibil la adresa: <https://ojs.aaai.org/index.php/ICWSM/article/view/14550/14399>.
3. Pang, Bo, și Lillian Lee. "Opinion Mining and Sentiment Analysis." *Foundations and Trends in Information Retrieval* 2.1-2 (2008): 1-135, Disponibil la adresa: <https://www.cs.cornell.edu/home/llee/omsa/omsa.pdf>.
4. Devlin, Jacob, et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *arXiv preprint arXiv:1810.04805* (2018), Disponibil la adresa: <https://arxiv.org/abs/1810.04805>.
5. Bollen, Johan, Huina Mao, și Xiaojun Zeng. "Twitter Mood Predicts the Stock Market." *Journal of Computational Science* 2.1 (2011): 1-8, Disponibil la adresa: <https://www.sciencedirect.com/science/article/abs/pii/S187775031100007X>.
6. Thelwall, Mike, et al. "Sentiment Strength Detection in Short Informal Text." *Journal of the American Society for Information Science and Technology* 61.12 (2010): 2544-2558, Disponibil la adresa: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21416>.
7. Pang, Bo, Lillian Lee, și Shivakumar Vaithyanathan. "Thumbs up?: Sentiment Classification Using Machine Learning Techniques." *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Volume 10* (2002).
8. Paul Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992.
9. Seo, Yeong-Seok & Huh, Jun-Ho. (2019). Automatic Emotion-Based Music Classification for Supporting Intelligent IoT Applications. Electronics. 8. 164. 10.3390/electronics8020164.

# Anexe

## Anexa1 – Detalii elemente teoretice

..

## Anexa2 – Codul sursă

Elementele de cod în cadrul anexelor trebuiesc scrise în Courier New 10, fie cu instrucţiunile **boldate**, fie cu colorarea elementelor limbajului etc.

Exemplu:

**#include <stdio.h>**

**#include <conio.h>**

**int main()**

**{**

**printf("Hello, World!\n");**

**getch();**

**return 0;**

**}**