

Impactul utilizarii retelelor de socializare asupra mintii

Tema 31

Neaga Matei-Dragos

Țîrlea Diana-Maria

Dinu Marius-Valentin

Grupa 543B

Cuprins

1. Tematica	3
Ce este acest studiu?	3
2. Introducere / Context	3
De ce este relevantă această cercetare?	3
Contextul Actual	3
Schimbările Sociale	3
Impactul Tehnologic	4
Interacțiunea Între Tehnologie și Sănătatea Mintală	4
Necesitatea Investigării	4
Machine learning	4
3. Stadiul actual	6
4. Descriere problema	9
Flowchart	10
5. Descrierea setului de date furnizat:	10
6. Prelucrarea datelor	12
6.2 Rezultate rulare cod	12
7. Interpretarea rezultatelor	16
Anexe	18
Bibliografie	26

1. Tematica

Scopul principal al acestei lucrări este de a atribui un calificativ care să reflecte gradul de dependență al indivizilor față de rețelele de socializare, utilizând un set de date complex și diversificat. Prin antrenarea unui program de machine learning pe acest set de date, se intenționează obținerea unei evaluări obiective și precise a nivelului de adicție la social media, oferind astfel o înțelegere mai profundă a impactului acestei dependențe asupra sănătății mintale.

Ce este acest studiu?

Acest studiu reprezintă o investigație amănunțită asupra impactului rețelelor de socializare asupra indivizilor, având în vedere complexitatea interacțiunilor sociale, schimbările comportamentale și implicațiile pentru sănătatea mintală.

2. Introducere / Context

De ce este relevantă această cercetare?

Contextul Actual

Suntem martorii unei transformări continue a societății noastre către o lume digitalizată, în care rețelele de socializare au devenit componente esențiale ale vieții noastre cotidiene. Această interconectare digitală a avut un impact semnificativ asupra modului în care comunicăm, relaționăm și, implicit, asupra sănătății noastre mintale.

Schimbările Sociale

Schimbările sociale aduse de expansiunea rețelelor de socializare sunt evidente în modalitatea în care oamenii își construiesc și își prezintă identitățile, înțeleg conceptele de

prietenie și comunitate, și gestionează relațiile interpersonale. Este esențial să înțelegem cum aceste schimbări influențează starea noastră de bine și sănătatea mintală.

Impactul Tehnologic

Avansul tehnologic rapid și accesul facil la platformele de social media au schimbat dinamica interacțiunii sociale. Acest aspect ridică întrebări cu privire la modul în care utilizarea intensivă a acestor platforme poate afecta echilibrul mental, emoțional și relațional al individului.

Interacțiunea Între Tehnologie și Sănătatea Mintală

Explorarea acestei interacțiuni este crucială pentru a ne adapta la mediul digital în continuă evoluție și pentru a dezvolta strategii eficiente de menținere a sănătății mintale în era rețelelor de socializare. Înțelegerea impactului asupra stării noastre psihologice este esențială pentru a oferi intervenții eficiente și pentru a promova utilizarea sănătoasă a tehnologiei.

Necesitatea Investigării

Prin această cercetare, ne propunem să analizăm în profunzime cum utilizarea rețelelor de socializare poate afecta sănătatea mintală, deoarece aceasta este o preocupare crescută în rândul societății contemporane. Identificarea și înțelegerea factorilor de influență sunt etape cruciale pentru a dezvolta strategii de gestionare a impactului negativ și pentru a promova utilizarea conștientă și sănătoasă a rețelelor de socializare.

Machine learning

Machine learning reprezintă o ramură a inteligenței artificiale (IA) care se concentrează pe dezvoltarea algoritmilor și modelelor capabile să învețe din date și să facă previziuni sau să ia decizii fără a fi programate explicit. În contextul impactului utilizării rețelelor de socializare asupra sănătății mintale, machine learning poate fi implicat în mai multe moduri:

Analiza Sentimentului

Algoritmii de machine learning pot analiza postările și comentariile de pe rețelele de socializare pentru a evalua sentimentele exprimate. Aceasta poate oferi o înțelegere a modului în care utilizatorii percep și reacționează la diverse conținuturi.

Identificarea Factorilor de Stres

Machine learning poate fi folosit pentru a identifica modele sau factori care ar putea contribui la stresul sau anxietatea utilizatorilor. Analizând comportamentul online, algoritmiile pot evidenția momente sau situații care ar putea avea un impact negativ asupra sănătății mintale.

Personalizarea Conținutului

Algoritmiile pot fi utilizate pentru a personaliza conținutul afișat pe rețelele de socializare, astfel încât să se evite subiecte sau tipuri de conținut care au fost asociate în trecut cu reacții negative ale utilizatorilor.

Sisteme de Suport și Intervenție

Machine learning poate contribui la dezvoltarea sistemelor de susținere și intervenție, detectând semne ale unor probleme de sănătate mintală și oferind sugestii sau resurse relevante pentru a ajuta utilizatorii.

Analiza Părerilor și Reacțiilor Colective:

Algoritmiile pot evalua reacțiile colective la anumite evenimente sau subiecte, furnizând o înțelegere mai amplă a modului în care anumite aspecte ale conținutului online pot afecta comunitățile.

Este important să menționăm că utilizarea machine learning în acest context implică, de asemenea, provocări etice și de confidențialitate, precum și necesitatea de a lua în considerare aspecte precum interpretarea corectă a datelor și evitarea discriminării. Abordările responsabile și etice în dezvoltarea și implementarea acestor tehnologii sunt esențiale pentru a asigura beneficii reale și durabile pentru sănătatea mintală a utilizatorilor.

3. Stadiul actual

1. “Using Social Media for Social Comparison and Feedback-Seeking: Gender and Popularity Moderate Associations with Depressive Symptoms” - Jacqueline Nesi & Mitchell J. Prinstein (2015)

Acest studiu a investigat comportamente tehnologice specifice (comparație socială și căutarea de feedback interpersonal) care pot interacționa cu caracteristicile individuale offline pentru a prezice simptomele concomitente de depresie la adolescenți. Un total de 619 de elevi (57% fete; vârsta medie 14,6 ani) au completat chestionare auto-raportate la două momente diferite. Adolescenții au raportat nivelurile de simptome depresive la început, iar un an mai târziu au raportat simptomele depresive, frecvența utilizării tehnologiei (telefoane mobile, Facebook și Instagram), căutarea excesivă de reasigurare și comparația socială și căutarea de feedback bazate pe tehnologie. Adolescenții au completat, de asemenea, nominalizări sociometrice ale popularității. Conform ipotezelor, comparația socială și căutarea de feedback bazate pe tehnologie au fost asociate cu simptomele depresive. Popularitatea și genul au acționat ca moderatorii ai acestui efect, astfel încât asocierea a fost deosebit de puternică în rândul femeilor și al adolescenților cu popularitate scăzută. Asociațiile au fost găsite dincolo de efectele frecvenței generale a utilizării tehnologiei, a căutării excesive offline de reasigurare și a simptomelor depresive anterioare.

Metodele utilizate în acest studiu au implicat colectarea datelor prin intermediul chestionarelor auto-raportate completate de către 619 de elevi la două momente diferite. La începutul studiului, adolescenții au raportat nivelurile lor de simptome depresive, iar un an mai târziu au furnizat informații despre simptomele depresive, frecvența utilizării tehnologiei (telefoane mobile, Facebook și Instagram), căutarea excesivă de reasigurare și comportamentele specifice de comparație socială și căutare de feedback bazate pe tehnologie. Pentru a evalua popularitatea, adolescenții au completat nominalizări sociometrice.

Analiza datelor a implicat metode statistice avansate, inclusiv analize de regresie, pentru a evalua asocierea dintre comportamentele tehnologice specifice și simptomele depresive. Au fost folosite moderări pentru a examina rolul popularității și al genului în interacțiunea cu aceste

asociere. De asemenea, au fost controlate variabilele confundatoare, precum frecvența generală a utilizării tehnologiei, căutarea excesivă offline de reasigurare și simptomele depresive anterioare.

Aceste metode au permis cercetătorilor să identifice și să evalueze impactul specific al comportamentelor tehnologice asupra sănătății mentale a adolescenților, luând în considerare și factori moderatori importanți.

Concluziile subliniază utilitatea examinării consecințelor psihologice ale utilizării tehnologiei de către adolescenți în cadrul modelelor interpersonale existente ale depresiei la adolescenți și sugerează importanța unor abordări mai nuanțate în studiul utilizării media a adolescenților. [\[1\]](#)

2. “The associations between aggressive behaviors and internet addiction and online activities in adolescents” - Chih-Hung Ko, Ju-Yu Yen, Shu-Chun Liu, Chi-Fen Huang, Cheng-Fang Yen (2009)

Acest studiu își propune să evalueze:

- asocierea dintre dependența de internet și comportamentele agresive, precum și efectele moderate ale genului, școlii și depresiei asupra acestei asocieri;
- legătura dintre activitățile pe internet și comportamentele agresive.

Metode: În cadrul acestui studiu, au fost recrutați 9405 adolescenți care au completat chestionarele. Au fost evaluate comportamentele agresive, cu sau fără dependența de internet, activitățile pe internet, datele demografice, cu sau fără depresie, autoestima, funcția familială și vizionarea programelor TV violente.

Rezultate: Rezultatele au arătat că, după controlul pentru efectele factorilor asociați și vizionarea programelor TV violente, adolescenții cu dependență de internet aveau o probabilitate mai mare de a manifesta comportamente agresive în anul precedent. Asocierea era mai

semnificativă în rândul adolescenților de gimnaziu decât în rândul celor de liceu/școală profesională. Conversațiile online, vizionarea conținutului pentru adulți pe internet, jocurile online, jocurile de noroc online și sistemul de bord au fost toate asociate cu comportamentele agresive.

Concluzii: Rezultatele sugerează că programele preventive pentru comportamentele agresive ar trebui să acorde atenție dependenței de internet în rândul adolescenților. De asemenea, intervenția pentru prevenirea efectelor dependenței de internet asupra comportamentelor agresive ar trebui să fie realizată cât mai devreme posibil. [\[2\]](#)

3. "Seeing everyone else's highlight reels: How Facebook usage is linked to depressive symptoms" - Mai-Ly Nguyen Steers, Robert E. Wickham, Linda K. Acitelli (2014)

Scopul cercetării este de a analiza utilizările Facebook și efectele psihologice asociate acestora, având în vedere conceptul psihanalitic de "containment" (conținție). Am integrat acest concept cu formele anxioase și evitante de atașament pentru a crea un cadru de investigare a utilizărilor active și pasive ale Facebook. Prin intermediul unui chestionar administrat la 275 de studenți israelieni, am constatat că nevoia de "containment" a prezis conectarea și postarea pe Facebook după experiențe negative. În plus, îndeplinirea nevoii de "containment" a fost asociată cu postarea pe Facebook. Am observat, de asemenea, o schimbare pozitivă în sentimentele despre sine și în sensul subiectiv al vizibilității după postarea pe Facebook. Aceste rezultate aduc o abordare nouă pentru explorarea atât a utilizării Facebook, cât și a efectelor sale.

Metode: Cercetarea s-a desfășurat printr-un chestionar administrat la 275 de studenți israelieni, utilizând conceptul psihanalitic de "containment" și formele de atașament anxioase și evitante. Chestionarul a fost conceput pentru a investiga modul în care nevoia de "containment" influențează comportamentele de conectare și postare pe Facebook după experiențe negative.

Concluzii: Rezultatele oferă o perspectivă nouă asupra utilizării Facebook și a efectelor sale, evidențiind importanța nevoii de "containment" în relație cu comportamentele specifice pe platformă și impactul pozitiv asupra stării de bine și a percepției de vizibilitate a individului. [\[3\]](#)

4. “Social media use and depression in adolescents: a scoping review” - Carol Vidal , Tenzin Lhaksampa , Leslie Miller , Rheanna Platt (2020)

Scopul acestei cercetări este de a examina asociația bidirecțională dintre utilizarea rețelelor de socializare (RS), în special a platformelor de socializare (PS), și depresie și tendințe suicidale în rândul adolescenților, în contextul creșterii semnificative a depresiei și comportamentului suicidar în ultimele două decenii, concomitent cu popularizarea mediilor de socializare digitală, utilizate extensiv de către adolescenți. Prin analiza tematică a studiilor revizuite, au fost identificate patru teme principale în utilizarea RS: cantitatea de utilizare, calitatea utilizării, aspectele sociale asociate cu utilizarea RS și dezvăluirea simptomelor de sănătate mintală. În avansarea cercetării în acest domeniu, ar fi benefică implementarea unor designuri longitudinale, utilizarea unor măsurători obiective și actuale ale utilizării RS, cercetarea mecanismelor asociate între utilizarea RS și depresie și tendințe suicidale, precum și cercetarea în populații clinice pentru a ghida practica clinică.

Metode: Cercetarea a fost realizată prin analiza tematică a studiilor care explorează asociația dintre utilizarea rețelelor de socializare, în special platformele de socializare, și depresie și tendințe suicidale în rândul adolescenților.

Concluzii: Această revizuire evidențiază necesitatea unor abordări longitudinale, măsuri obiective și actuale ale utilizării rețelelor de socializare, investigarea mecanismelor subiacente asocierii cu depresia și tendințele suicidale și extinderea cercetării în populații clinice pentru a furniza informații relevante pentru practica clinică. [\[4\]](#)

4. Descriere problema

Problematica temei se concentrează pe analiza dependenței de rețelele sociale și evaluarea impactului acestora asupra unor factori cheie, inclusiv calitatea somnului și starea de depresie.

Flowchart

Are drept obiect stabilirea pașilor care trebuie urmați

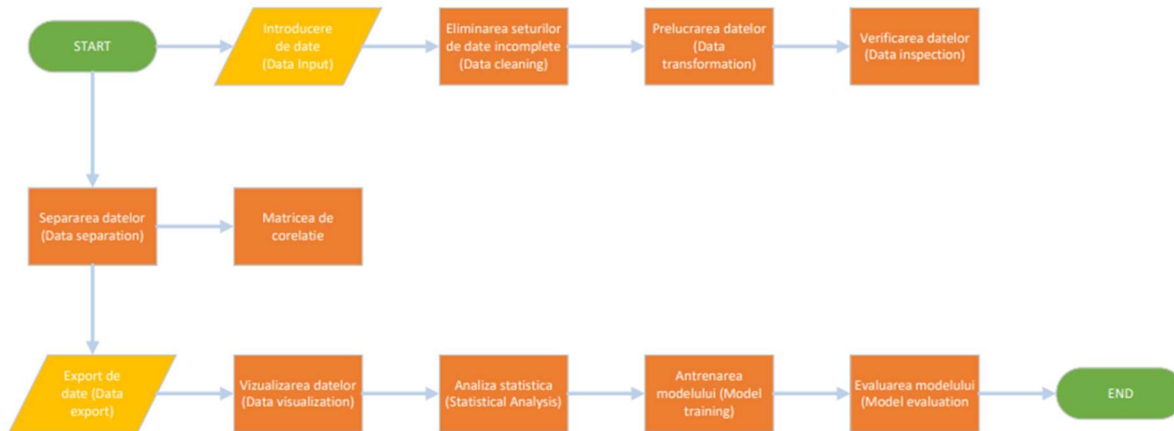


Figura 4.1 – flowchart-ul respectat pentru rezolvarea problemei propuse

5. Descrierea setului de date furnizat:

5.1 Număr de Eșantioane

- 481 persoane

5.2 Caracteristici:

- Demografice: Vârstă, Gen, Stare civilă, Statut ocupational, Afilieri la organizații.
- Utilizarea rețelelor de socializare: Frecvența utilizării, Platforme preferate, Timpul mediu zilnic petrecut.
- Comportament pe rețelele de socializare: Utilizarea fără scop, Distragerea, Agitația resimțită, Distractibilitate, Deranjul cauzat de îngrijorări.

- Aspecte emoționale și mentale: Dificultăți de concentrare, Comparare cu persoane de succes, Reacția emoțională la comparații, Căutarea validării, Frecvența stărilor depresive, Fluctuații în interesul pentru activități zilnice, Problemele de somn.

5.3. Etichete:

Setul de date poate include etichete care să indice categorii sau rezultate relevante pentru fiecare participant în funcție de răspunsurile lor.

5.4. Unități de Măsură și Intervale:

- Vârstă: Măsurată în ani.
- Timpul petrecut pe social media: Măsurat în ore sau minute.
- Scoruri de la întrebările cu scale: Măsurate pe o scară de la 1 la 5.

5.5. Semnificație:

- Setul de date furnizează informații esențiale despre comportamentul, starea emoțională și caracteristicile demografice ale participanților în contextul utilizării rețelelor de socializare.

- Scopul este de a investiga și înțelege impactul utilizării intensive a rețelelor de socializare asupra sănătății mintale și a altor aspecte legate de comportamentul online.

Analiza acestui set de date poate oferi perspective semnificative în ceea ce privește relațiile dintre diferite variabile și să ofere informații relevante pentru cercetările privind impactul social media asupra sănătății mintale.

6. Prelucrarea datelor

S-a optat pentru învățarea supervizată de tip regresie, intrucât studiul se bazează pe probleme privind calitatea somnului (coloana 20 din setul de date tip CSV), aceasta fiind dependentă de orele petrecute pe rețelele sociale, adică de intrarea modelului. Astfel se încadrează perfect în definiția învățării supervizate de tip regresie.

Se exportează matricea de corelație 16x16 între toate caracteristicile din setul de date, mai puțin ieșirea (calitatea somnului, coloana 20) într-un fișier tip Excel separat pentru o mai bună vizualizare a sa.

6.2 Rezultate rulare cod

Datele sunt uniforme

Se exportează matricea de corelație într-un fișier tip Excel separat pentru o mai bună vizualizare a sa

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	1	0.015869	-0.04691	-0.07735	0.545105	-0.36176	-0.22519	-0.23206	-0.12574	-0.22444	-0.2887	-0.26453	-0.09254	0.018018	-0.039	-0.30367	-0.24145
1	0.015869	1	0.009627	-0.03806	-0.061	0.083491	0.017895	0.009929	0.095135	0.082123	0.028237	0.04235	0.014551	0.042304	0.02525	0.042485	-0.00267
2	-0.04691	0.009627	1	0.050997	0.010686	0.111795	0.08823	0.13854	0.100205	0.113511	0.116952	0.11256	0.103176	0.008167	0.092497	0.116108	0.073866
3	-0.07735	-0.03806	0.050997	1	-0.09055	0.070648	0.023944	0.046051	0.132659	0.09007	0.138074	0.113376	0.077079	0.02383	0.044837	0.071542	0.043912
4	0.545105	-0.061	0.010686	-0.09055	1	-0.25726	-0.10988	-0.16945	-0.06028	-0.16721	-0.25548	-0.2072	-0.1092	0.007818	-0.03963	-0.27754	-0.167
5	-0.36176	0.083491	0.111795	0.070648	-0.25726	1	0.393532	0.380818	0.400058	0.342524	0.326245	0.321685	0.20061	0.050989	0.109225	0.330489	0.292408
6	-0.22519	0.017895	0.08823	0.023944	-0.10988	0.393532	1	0.463353	0.359031	0.365491	0.252082	0.288829	0.148203	0.087645	0.144519	0.315478	0.359379
7	-0.23206	0.009929	0.13854	0.046051	-0.16945	0.380818	0.463353	1	0.513354	0.607419	0.343307	0.527745	0.288504	0.078065	0.270144	0.379444	0.389132
8	-0.12574	0.095135	0.100205	0.132659	-0.06028	0.400058	0.359031	0.513354	1	0.514801	0.347036	0.46944	0.320973	0.134191	0.245925	0.328969	0.38477
9	-0.22444	0.082123	0.113511	0.09007	-0.16721	0.342524	0.365491	0.607419	0.514801	1	0.462673	0.662768	0.28612	0.01463	0.172585	0.407667	0.479993
10	-0.2887	0.028237	0.116952	0.138074	-0.25548	0.326245	0.252082	0.343307	0.347036	0.462673	1	0.545904	0.414301	0.020589	0.208847	0.588757	0.430348
11	-0.26453	0.04235	0.11256	0.113376	-0.2072	0.321685	0.288829	0.527745	0.46944	0.662768	0.545904	1	0.371156	0.060742	0.242035	0.509794	0.503196
12	-0.09254	0.014551	0.103176	0.077079	-0.1092	0.20061	0.148203	0.288504	0.320973	0.28612	0.414301	0.371156	1	-0.01147	0.417328	0.417801	0.374242
13	0.018018	0.042304	0.008167	0.02383	0.007818	0.050989	0.087645	0.078065	0.134191	0.01463	0.020589	0.060742	-0.01147	1	0.153574	0.03547	0.046153
14	-0.039	0.02525	0.092497	0.044837	-0.03963	0.109225	0.144519	0.270144	0.245925	0.172585	0.208847	0.242035	0.417328	0.153574	1	0.272125	0.278575
15	-0.30367	0.042485	0.116108	0.071542	-0.27754	0.330489	0.315478	0.379444	0.328969	0.407667	0.588757	0.509794	0.417801	0.03547	0.272125	1	0.494976
16	-0.24145	-0.00267	0.073866	0.043912	-0.167	0.292408	0.359379	0.389132	0.38477	0.479993	0.430348	0.503196	0.374242	0.046153	0.278575	0.494976	1

Figura 6.1 – matricea de corelație

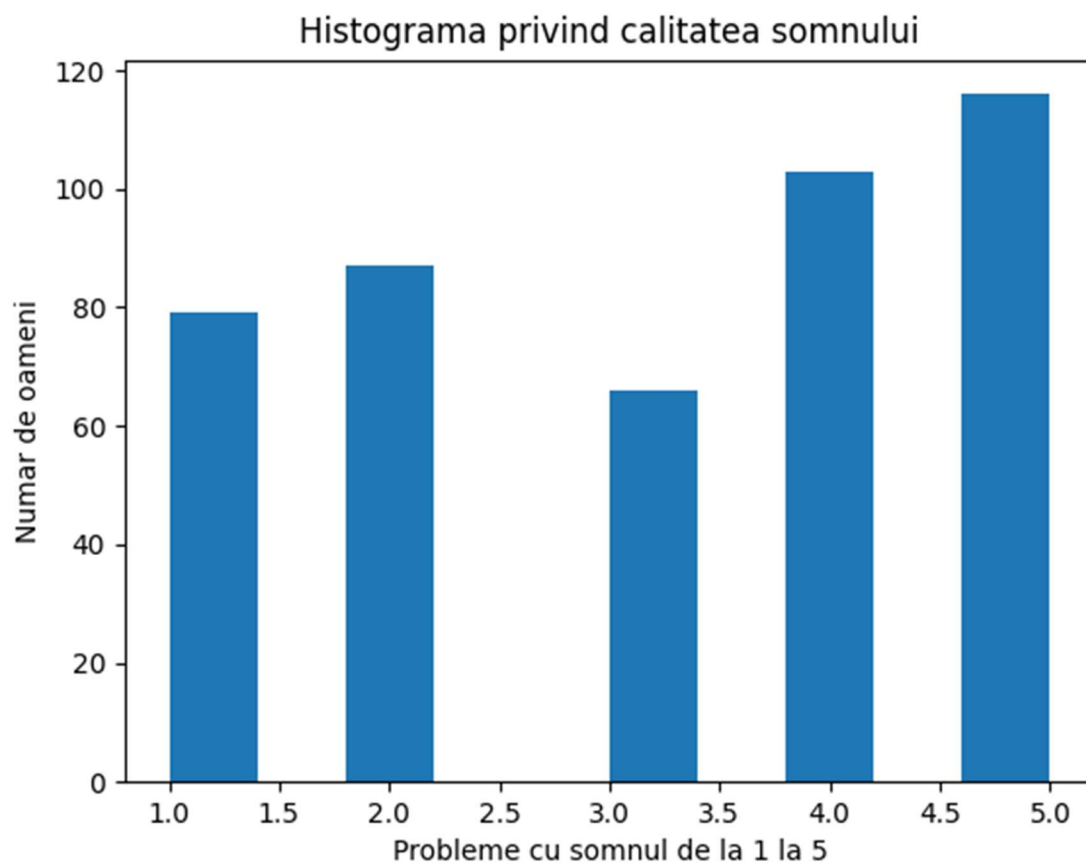


Figura 6.1 – histograma privind calitatea somnului

Deviatia standard a vectorului de valori corespunzatoare calitatii somnului este: 1.4616191574177457

Deviatia standard a vectorului de valori corespunzatoare timpului petrecut pe social media este:
1.5903190534023957

Media calitatii somnului, conform notelor oferite este: 3.2016632016632016

Media numarului de ore petrecut pe retelele sociale, conform notelor oferite este: 2.9085239085239087

Valoarea t: -2.9764598297994422

Valoarea p: 0.0029891232018100984

Gradele de libertate ale primului vector de valori, corespunzator cu timpul petrecut pe social media este: 481

Gradele de libertate ale celui de-al doilea vector de valori, corespunzator cu calitatea somnului este: 481

Grafic 2D density pentru o mai buna vizualizare a iesirii in functie de intrare, intrucat o mare majoritate a notelor acordate se suprapun, drept urmare dintr-un grafic tip scatter nu se pot trage concluzii pertinente.

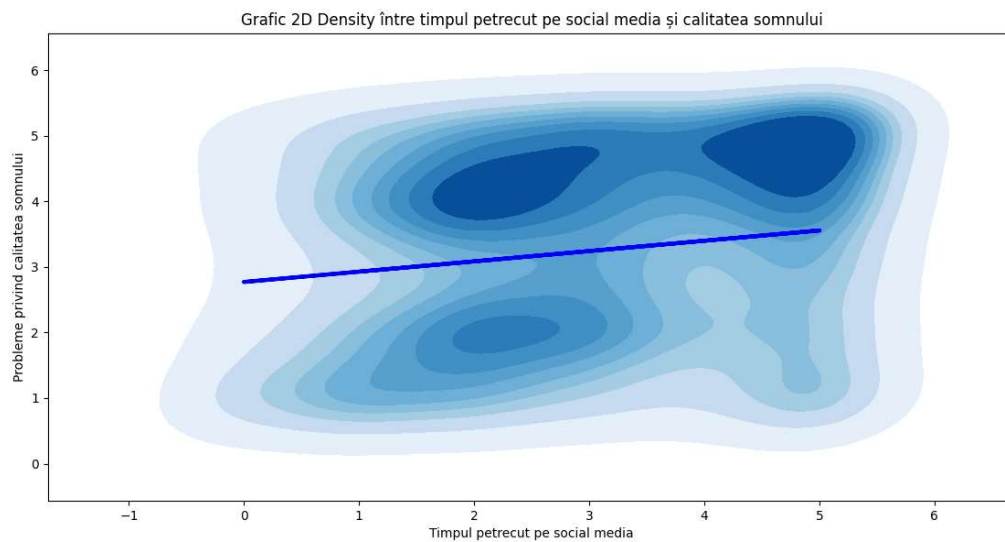


Figura 6.2 – graphic 2D privind calitatea somnului in functie de numarul de ore petrecut pe social media

Pentru a confirma cele afirmate anterior, se afiseaza un grafic tip scatter

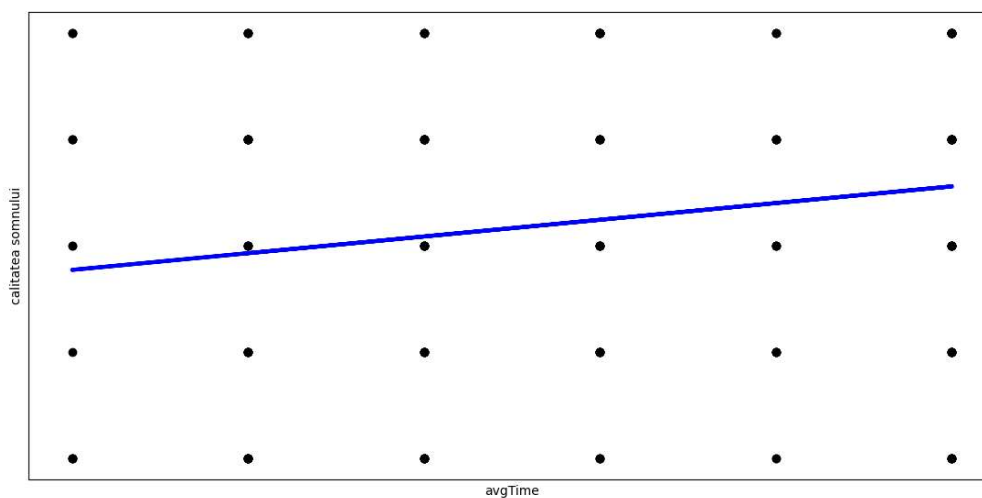


Figura 6.3 – graphic tip scatter privind calitatea somnului in functie de numarul de ore petrecut pe social media

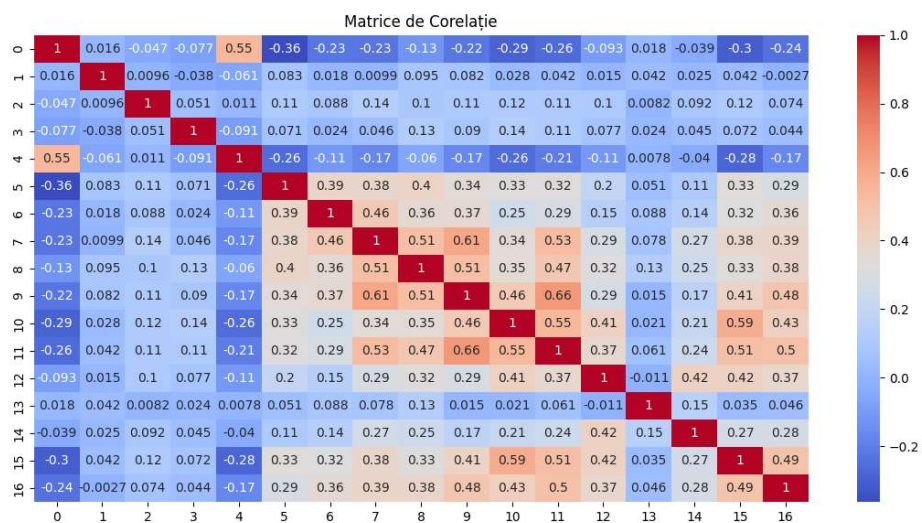


Figura 6.4 – heatmap al matricei de corelație

Invatare supervizata de tip regresie - rezultate

Coeficient de regresie:

[0.1570208]

Eroarea mediei patratice: 1.95

Coeficientul de determinare: 0.03

Invatare folosind algoritmul gaussian Naive Bayes - rezultate

Acuratetea modelului: 0.28

Precizia modelului: 0.18

7. Interpretarea rezultatelor

Pe baza rezultatelor furnizate, se pot trage următoarele concluzii, din codul atașat în anexa:

- Datele sunt uniforme, deci setul de date este echilibrat / nu există diferențe semnificative între variabilele măsurate.
- Matricea de corelație oferă informații despre relațiile liniare între diferitele caracteristici ale setului de date.
- Deviația standard a vectorului de valori pentru calitatea somnului și timpul petrecut pe social media oferă o măsură a dispersiei datelor în jurul mediei.
- Media calității somnului este în jur de 3.20, iar media timpului petrecut pe rețelele sociale este în jur de 2.91.
- În urma testului T efectuat între numărul de ore petrecute pe social media și calitatea somnului, Valorile negative pentru t și valoarea mică a p indică faptul că există o diferență semnificativă între cele două grupuri

- Se utilizează regresia liniară pentru a modela relația dintre variabila de intrare (numărul de ore petrecute pe rețele sociale) și variabila de ieșire (calitatea somnului), de unde rezulta: coeficient de regresie mic (aproximativ 0.16), iar eroarea medie pătrată este ridicată (aproximativ 1.95), indicând o ajustare slabă a modelului la date.

- Interpretarea subliniata profund de graficul 2D density este corelarea între numărul de ore petrecut pe rețelele sociale (5) cu cea mai slabă calitate a somnului (5), graficul de tip scatter neoferind informații concludente despre nicio corelare

- Se folosește algoritmul Gaussian Naive Bayes pentru clasificare, iar acuratețea modelului este de aproximativ 0.28 și precizia este de aproximativ 0.18, ceea ce indică o performanță modestă a modelului în clasificarea datelor

În concluzie, rezultatele sugerează că modelul de regresie liniar nu este cel mai optim pentru analiza datelor, iar modelul Naive Bayes are performanțe modeste în clasificarea datelor. Prin intermediul graficului 2D density se face o corelare bine evidentiată în acesta între numărul de ore petrecut pe rețelele de socializare și calitatea somnului.

Anexe

```
import pandas as pd
import numpy as np
import openpyxl
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from scipy import stats
from sklearn import linear_model, datasets
from sklearn.metrics import accuracy_score, precision_score, mean_squared_error, r2_score

df = pd.read_csv(r"C:\Users\MATEI NEAGA\Desktop\Inteligența Artificială\Proiect\setDate.csv",
                 encoding="ISO-8859-1")

def map_gender_to_numeric(gender):
    if gender.lower() == 'male':
        return 1
    elif gender.lower() == 'female':
        return 1
    else:
        return 5

# se aplica funcția pe coloana 'Gender'
gender = df['2. Gender'].apply(map_gender_to_numeric)

status = df["3. Relationship Status"].map({
    "Single": 0,
    "Divorced": 1,
    "In a relationship": 3,
    "Married": 5
}).astype('int64')

social_media = df["7. What social media platforms do you commonly use?"]

useSocials = df['6. Do you use social media?'].map({
    'No': 1,
    'Yes': 2
}).astype('int64')
```

```

def map_social_media_count(count):
    if count <= 2:
        return 1
    elif count <= 3:
        return 2
    elif count <= 4:
        return 3
    elif count <= 5:
        return 4
    elif count <= 6:
        return 5

# Count the number of social media platforms for each row
social_media_count = social_media.apply(lambda x: len(x.split(', ')))

# Map the count to a score using the custom function
socials = social_media_count.map(
    map_social_media_count).fillna(0).astype('int64')

# etichetare, se foloseste functia map pentru a inlocui valorile de tip text cu
# valori numerice, pentru a se putea face corelari ulterioare
avgTime = df["8. What is the average time you spend on social media every
day?"].map({
    "Less than an Hour": 0,
    "Between 1 and 2 hours": 1,
    "Between 2 and 3 hours": 2,
    "Between 3 and 4 hours": 3,
    "Between 4 and 5 hours": 4,
    "More than 5 hours": 5
}).astype('int64')

data_frames = [pd.to_numeric(
    df["1. What is your age?"].apply(lambda x: int(x))),
    pd.to_numeric(
        df["9. How often do you find yourself using Social media without a
specific purpose?"]),
    pd.to_numeric(
        df["10. How often do you get distracted by Social media when you are busy
doing something?"]),
    pd.to_numeric(
        df["11. Do you feel restless if you haven't used Social media in a while?"]),
    pd.to_numeric(
        df["12. On a scale of 1 to 5, how easily distracted are you?"]),

```

```

    pd.to_numeric(
    df["13. On a scale of 1 to 5, how much are you bothered by worries?"]),
    pd.to_numeric(
    df["14. Do you find it difficult to concentrate on things?"]),
    pd.to_numeric(
    df["15. On a scale of 1-5, how often do you compare yourself to other
successful people through the use of social media?"]),
    pd.to_numeric(
    df["16. Following the previous question, how do you feel about these
comparisons, generally speaking?"]),
    pd.to_numeric(
    df["17. How often do you look to seek validation from features of social
media?"]),
    pd.to_numeric(
    df["18. How often do you feel depressed or down?"]),
    pd.to_numeric(
    df["19. On a scale of 1 to 5, how frequently does your interest in daily
activities fluctuate?"]),
    pd.to_numeric(
    df["20. On a scale of 1 to 5, how often do you face issues regarding
sleep?"]),

]
# "1. What is your age?"
df_1 = data_frames[0]
# "9. How often do you find yourself using Social media without a specific
purpose?"
df_9 = data_frames[1]
# "10. How often do you get distracted by Social media when you are busy doing
something?"
df_10 = data_frames[2]
# "11. Do you feel restless if you haven't used Social media in a while?"
df_11 = data_frames[3]
# "12. On a scale of 1 to 5, how easily distracted are you?"
df_12 = data_frames[4]
# "13. On a scale of 1 to 5, how much are you bothered by worries?"
df_13 = data_frames[5]
# "14. Do you find it difficult to concentrate on things?"
df_14 = data_frames[6]
# "15. On a scale of 1-5, how often do you compare yourself to other successful
people through the use of social media?"
df_15 = data_frames[7]
# "16. Following the previous question, how do you feel about these comparisons,
generally speaking?"

```

```

df_16 = data_frames[8]
# "17. How often do you look to seek validation from features of social media?"
df_17 = data_frames[9]
# "18. How often do you feel depressed or down?"
df_18 = data_frames[10]
# "19. On a scale of 1 to 5, how frequently does your interest in daily
activities fluctuate?"
df_19 = data_frames[11]
# "20. On a scale of 1 to 5, how often do you face issues regarding sleep?"
iesire = data_frames[12]

# verificare uniformitate date
ok = 0
for data_frame in data_frames:
    if data_frame.dtype == 'int64':
        ok += 1
    else:
        ok -= 1

# se verifica uniformitatea datelor
if ok == len(data_frames):
    print('Datele sunt uniforme')
    # print(f'Matricea de corelatie: \n', matrice_corelatie)

print('Se exporteaza matricea de corelatie intr-un fisier tip Excel separat
pentru o mai buna vizualizare a sa')
print('\n')
# se creeaza un dataframe pentru matricea de corelatie
all_features = pd.concat([df_1, gender, useSocials, socials, status, avgTime,
df_9, df_10,
                        df_11, df_12, df_13, df_14, df_15, df_16, df_17, df_18,
df_19], axis=1)
# matricea de corelatie a celorlalte caracteristici, in afara de iesirea Y,
figurand ca lipsa somnului (coloana 20)
matrice_corelatie_completa = np.corrcoef(all_features, rowvar=False)
# se scrie matricea de corelatie intr-un excel separat
export_matrice_corelatie_completa = pd.DataFrame(
    data=matrice_corelatie_completa)
export_matrice_corelatie_completa.to_excel(
    r'C:\Users\MATEI NEAGA\Desktop\Inteligenta
Artificiala\Proiect\matrice_corelatie_completa.xlsx')

```

```

# Histograma privind calitatea somnului

plt.hist(iesire)
plt.xlabel('Probleme cu somnul de la 1 la 5')
plt.ylabel('Numar de oameni')
plt.title('Histograma privind calitatea somnului')
plt.show()

# test T intre numarul de ore petrecut pe social media si calitatea somnului

# deviatia standard
print("Deviatia standard a vectorului de valori corespunzatoare calitatii somnului este:", stats.tstd(iesire))
print("Deviatia standard a vectorului de valori corespunzatoare timpului petrecut pe social media este:", stats.tstd(avgTime))
print('\n')

# media calitatii somnului
print('Media calitatii somnului, conform notelor oferite este:', iesire.mean())
print('Media numarului de ore petrecut pe retelele sociale, conform notelor oferite este:', avgTime.mean())
print('\n')

# parametrii t si p
t, p = stats.ttest_ind(avgTime, iesire)
# intensitatea diferentei dintre grupuri
print("Valoarea t:", t)
# probabilitatea de a obtine o valoare t cel putin la fel de mare ca valoarea obtinuta anterior
print("Valoarea p:", p)
print('\n')

# numarul de valori din setul de date care pot varia pentru a putea face comparatii
print('Gradele de libertate ale primului vector de valori, corespunzator cu timpul petrecut pe social media este:', len(avgTime))
print('Gradele de libertate ale celui de-al doilea vector de valori, corespunzator cu calitatea somnului este:', len(iesire))
print('\n')

avgTime = avgTime.values.reshape(-1, 1)
iesire = iesire.values.reshape(-1, 1)
iesire = iesire.ravel()
# antrenarea si prezicerea modelului

```

```

# test_size 40% reprezinta procentul de date convertite in date de test, restul
fiind date de antrenament
# random_state = 42 reprezinta intensitatea cu care sunt amestecate datele
avgTime_train, avgTime_test, iesire_train, iesire_test = train_test_split(
    avgTime, iesire, test_size=0.4, random_state=42)

# invatare supervizata de tip Logistica regresie
print('Invatare supervizata de tip regresie')

# model de regresie liniara
regr = linear_model.LinearRegression()

# antrenarea modelului
regr.fit(avgTime_train, iesire_train)

# predictia modelului
iesire_pred = regr.predict(avgTime_test)

# The coefficients
print("Coeficient de regresie: \n", regr.coef_)
# The mean squared error
print("Eroarea mediei patratice: %.2f" %
      mean_squared_error(iesire_test, iesire_pred))
# The coefficient of determination: 1 is perfect prediction
print("Coeficientul de determinare: %.2f" %
      r2_score(iesire_test, iesire_pred))

# varianta 1

# grafic
plt.scatter(avgTime_test, iesire_test, color="black")
plt.plot(avgTime_test, iesire_pred, color="blue", linewidth=3)

plt.xticks(())
plt.yticks(())
plt.xlabel(('avgTime'))
plt.ylabel(('calitatea somnului'))

plt.show()

# varianta2

# creare un dataframe dintre coloana de intrare si coloana de iesire
dataframe_kde = pd.DataFrame(

```

```

    {'avgTime': avgTime.ravel(), 'calitatea_somnului': iesire})

# grafic 2D Density
print('Grafic 2D density pentru o mai buna vizualizare a iesirii in functie de
intrare, intrucat o mare majoritate a notelor acordate se suprapun, drept urmare
dintr-un grafic tip scatter nu se pot trage concluzii pertinente.')
sns.kdeplot(data=dataframe_kde, x='avgTime',
            y='calitatea_somnului', fill=True, cmap='Blues')

plt.plot(avgTime_test, iesire_pred, color="blue", linewidth=3)
plt.xlabel('Timpul petrecut pe social media')
plt.ylabel('Probleme privind calitatea somnului')
plt.title(
    'Grafic 2D Density între timpul petrecut pe social media și calitatea
somnului')
plt.show()

# varianta 3
# plt.scatter(avgTime, iesire, alpha=0.5) # Adjust alpha for transparency

# # Add labels and title
# plt.xlabel('X-axis')
# plt.ylabel('Y-axis')
# plt.title('Scatter Plot with Overlapping Values')

# Show the plot
# plt.show()

# varianta 4
# sns.scatterplot(x='avgTime', y='calitatea_somnului', data=dataframe_kde,
alpha=0.5,
#                 color='blue', label='Dataset 1')

# # Add linear trendline
# sns.regplot(x='avgTime', y='calitatea_somnului',
#             data=dataframe_kde, scatter=False, color='blue')

# # Set labels and title
# plt.xlabel('X-axis')
# plt.ylabel('Y-axis')
# plt.title('Scatter Plot with Overlapping Values and Trendlines')

# # Show the plot

```



```

# plt.legend()
# plt.show()

# # Matrice de corelație tip HeatMap
sns.heatmap(matrice_corelatie_completa, annot=True,
            cmap='coolwarm')
plt.title('Matrice de Corelație')
plt.show()

# Algoritm Gaussian Naive Bayes

# Pregatirea datelor pentru antrenare si testare
X = all_features.values
y = iesire

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.4, random_state=42)

# Gaussian Naive Bayes
print('\n')
print('Invatare folosind algoritmul gaussian Naive Bayes')
gnb = GaussianNB()
gnb.fit(X_train, y_train)
y_pred = gnb.predict(X_test)

# Se evalueaza performanta
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average='weighted')

print(f'Acuratetea modelului: {accuracy:.2f}')
print(f'Precizia modelului: {precision:.2f}')

```

Bibliografie

- [1]. Nesi, J., & Prinstein, M. J. (2015). Using social media for social comparison and feedback-seeking: Gender and popularity moderate associations with depressive symptoms. *Journal of abnormal child psychology*, 43(8), 1427-1438.

- [2]. Ko, C. H., Yen, J. Y., Liu, S. C., Huang, C. F., & Yen, C. F. (2009). The associations between aggressive behaviors and Internet addiction and online activities in adolescents. *Journal of Adolescent Health*, 44(6), 598-605.

- [3]. Steers, M. L. N., Wickham, R. E., & Acitelli, L. K. (2014). Seeing everyone else's highlight reels: How Facebook usage is linked to depressive symptoms. *Journal of Social and Clinical Psychology*, 33(8), 701-731.

- [4]. Vidal, C., Lhaksampa, T., Miller, L., & Platt, R. (2020). Social media use and depression in adolescents: a scoping review. *International Review of Psychiatry*, 32(3), 235-253.