

Emotion Analysis

Dogăreci Bianca-Alexandra
Potlog Ioana
Negoită-Crețu Raluca
Spătaru Mara

Introducere

Scopul proiectului

Am realizat următoarele experimente în vederea clasificării emoțiilor:

- **Evaluare in-domain**
 - Dataset 1 (10 emoții)
 - Dataset 2 (6 emoții)
- **Generalizare cross-domain** (inclusiv evaluare filtrată)
 - antrenare pe Dataset 1 și testare pe Dataset 2
 - antrenare pe Dataset 2 și testare pe Dataset 1

Modele utilizate:

- Multinomial Naive Bayes
- Random Forest Classifier
- Logistic Regression
- Support Vector Machine (SVM)
- BERT

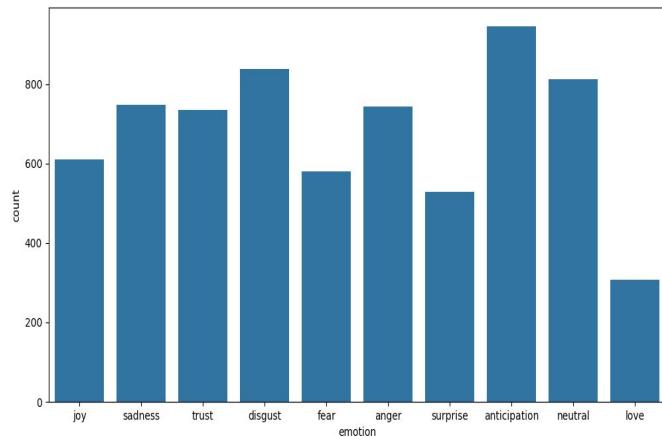
Setul de date

Sursa datelor

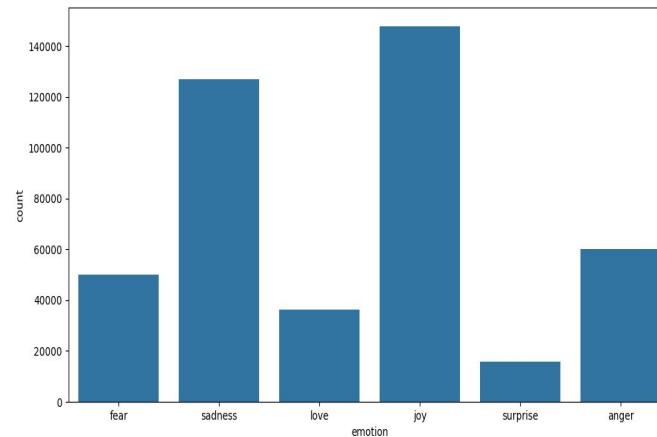
- **Dataset 1: uavster/Llama3_8b-emotion_multiclass-Plutchik** ([Hugging Face](#))
 - 6 539 exemple, format CSV
 - 9 emoții (8 conform Plutchik + neutral)
 - a fost adăugată manual cea de a 10-a emoție (Love) cu 300 exemple generate cu AI pentru a conține toate cele 6 emoții prezente în Dataset 2
- **Dataset 2: echung682/emotion-analysis-tweets** ([Hugging Face](#))
 - 436 809 exemple, format CSV
 - 6 emoții: anger, fear, joy, love, sadness, surprise
- ambele seturi de date au fost împărțite sub forma 70% antrenare și 30% testare

Setul de date

Frecvența categoriilor de emoții din Dataset 1



Frecvența categoriilor de emoții din Dataset 2



Preprocesarea datelor

Structura datelor

- Fiecare instanță este un text scurt (propoziție sau fragment de propoziție)
- Asociat cu o singură etichetă de emoție (ex: *joy*, *sadness*, *anger*, etc.)
- Emoțiile sunt exprimate în limba engleză
- Majoritatea textelor sunt scurte – acest aspect a fost confirmat prin analiza numărului de cuvinte după lematizare

Preprocesare aplicată

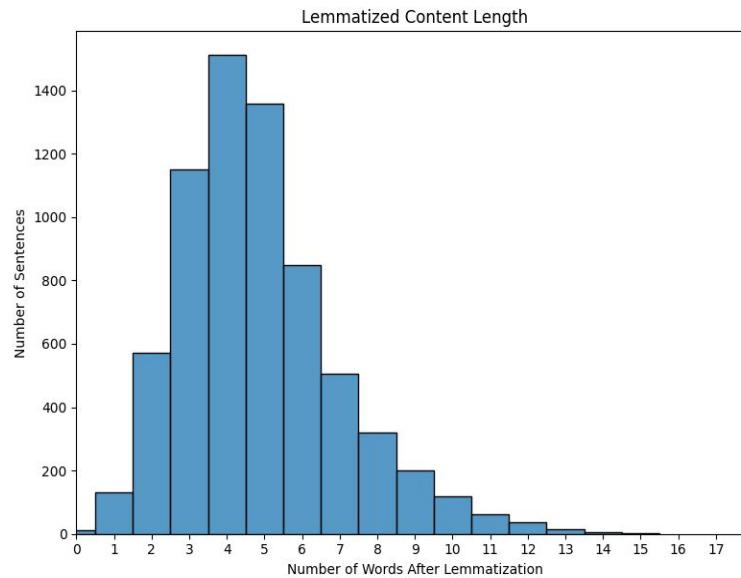
- Eliminarea semnelor de punctuație fără relevanță pentru clasificarea emoțiilor
- Înlocuirea emoji-urilor cu termeni descriptivi (ex: ":" → *sad*)
- Tokenizare – segmentarea textului în cuvinte individuale
- Eliminarea cuvintelor de tip *stop words* (ex: *the*, *is*, *are*) pentru reducerea zgomotului
- Lematizare – transformarea cuvintelor la forma lor de bază (ex: *tickets* → *ticket*)

Acstea etape au fost aplicate pentru a reduce zgomotul, a standardiza structura lexicală și a păstra doar informațiile relevante pentru clasificarea corectă a emoțiilor.

Preprocesarea datelor

Distribuția lungimii propozițiilor

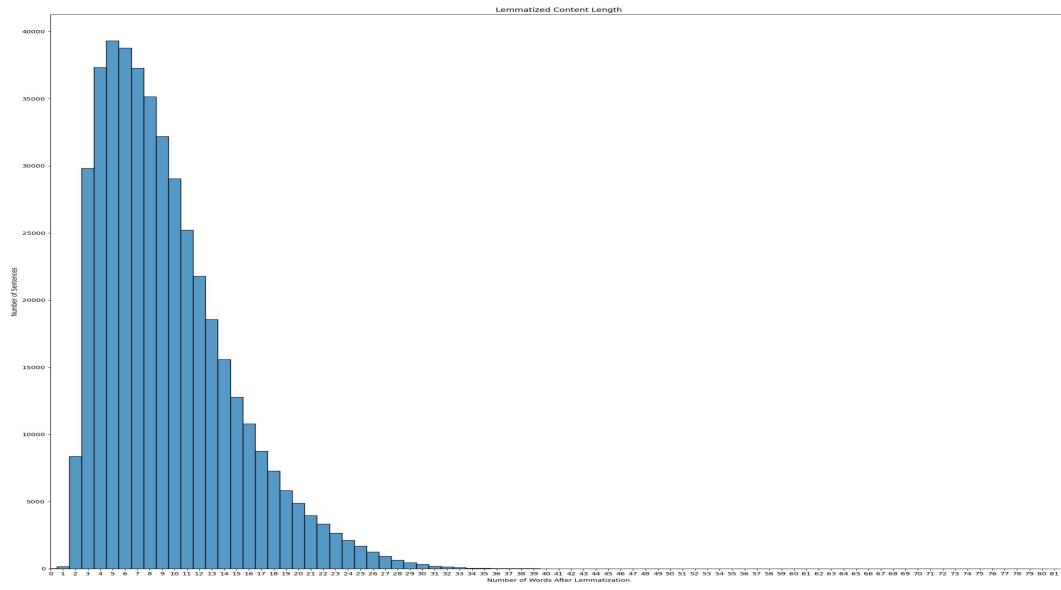
după lematizare – Dataset 1



Preprocesarea datelor

Distribuția lungimii propozițiilor

după lematizare – Dataset 2



Extractia caracteristicilor (Feature Extraction)

De ce e necesară?

După procesarea textului, acesta trebuie convertit într-o formă numerică fixă pentru a putea fi utilizat de modelele de învățare automată.

Tehnici utilizate în învățarea supervizată:

- **Count Occurrence** – contorizează aparițiile fiecărui cuvânt în propoziție, rezultând un vector de frecvențe
- **TF-IDF (Term Frequency – Inverse Document Frequency)** – ajustează frecvența în funcție de cât de comun este un cuvânt în toate documentele, acordând mai multă importanță celor distinctive și relevante

Limitări:

- Ambele metode generează reprezentări **sparse** și **lipsite de context**
- Nu surprind sensul cuvintelor sau relațiile dintre ele, ceea ce poate afecta acuratețea predicțiilor

Multinomial Naive Bayes

Descriere:

- Clasificator probabilistic bazat pe teorema lui Bayes
- Calculează probabilitatea ca un text să aparțină fiecărei clase de emoții
- Folosește frecvența cuvintelor pentru a estima probabilitățile
- Asumă independentă condițională între cuvinte (feature)

Caracteristici cheie:

- Model probabilistic simplu și eficient din punct de vedere computațional
- Funcționează bine cu texte reprezentate ca vectori de frecvență a cuvintelor
- Ignoră relațiile dintre cuvinte (asumția de independentă)
- Oferă un echilibru bun între eficiență și performanță predictivă

Multinomial Naive Bayes

Experiment 1: Model antrenat pe Dataset 1 și testat pe Dataset 1

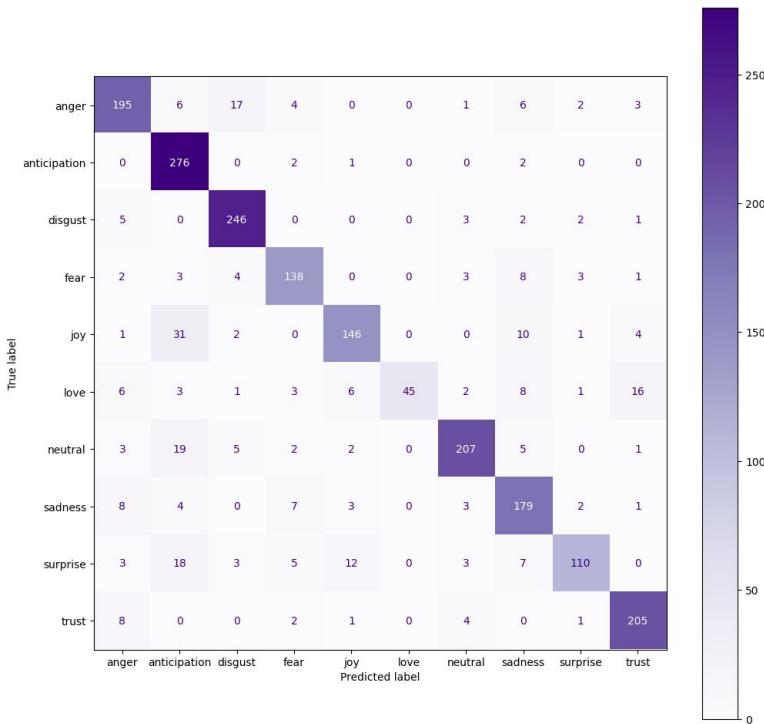
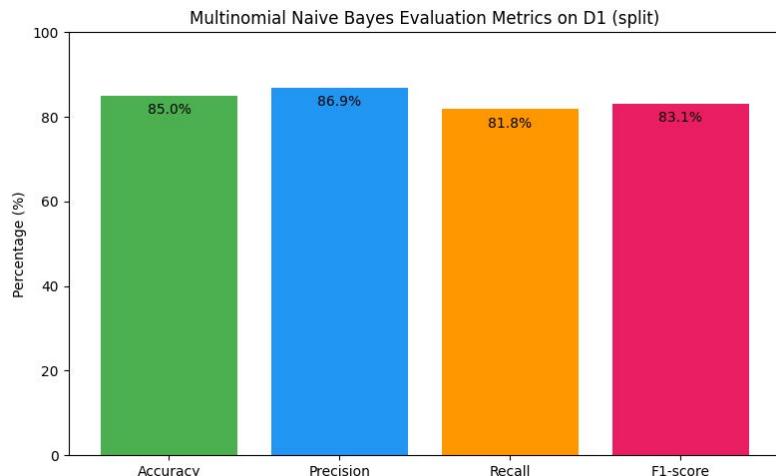
Performanță

- Acuratețe: **85.01%**
- Precizie: **86.86%**
- Recall: **81.84%**
- F1-Score: **83.13%**

Performanță pe clase de emoții

- **Cea mai bună performanță:** *disgust* ($F1: 0.91$), *trust* ($F1: 0.91$), *anticipation* ($F1: 0.86$)
 - Precizie și recall constant ridicate
- **Cea mai slabă performanță:** *love*
 - Precizie perfectă (1.00), dar recall scăzut: **0.49**
 - Modelul nu face niciodată predicții fals pozitive pentru *love*, dar ratează foarte multe instanțe: $F1= 0.66$

Multinomial Naive Bayes



Multinomial Naive Bayes

Experiment 2: Model antrenat pe Dataset 1 și testat pe Dataset 2 (toate emoțiile)

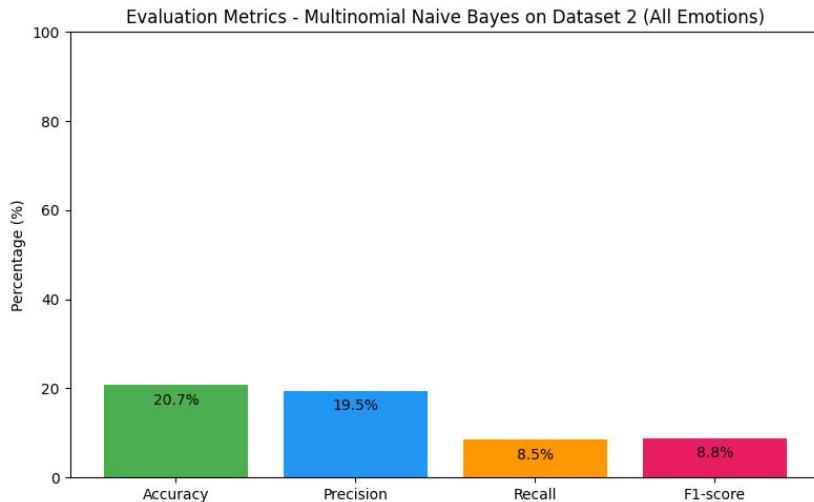
Performanță

- Acuratețe: **20.67%**
- Precizie: **19.46%**
- Recall: **8.45%**
- F1-Score: **8.72%**

Performanță pe clase de emoții

- Modelul arată o **generalizare extrem de slabă** pe Dataset 2
- Categoriile *anticipation, disgust, neutral, trust* complet nerecunoscute (nu există în setul de test)
- *Sadness* este cea mai bine recunoscută emoție (**F1: 0.42, recall: 0.57**)
- *Joy și fear* au performanță foarte slabă (F1: **0.09** și **0.16** respectiv)
- *Love și surprise* sunt aproape nerecunoscute (F1 <**0.07**)

Multinomial Naive Bayes



Multinomial Naive Bayes

Experiment 3: Model antrenat pe Dataset 1 și testat pe Dataset 2 (doar emoțiile comune dintre cele 2 dataset-uri)

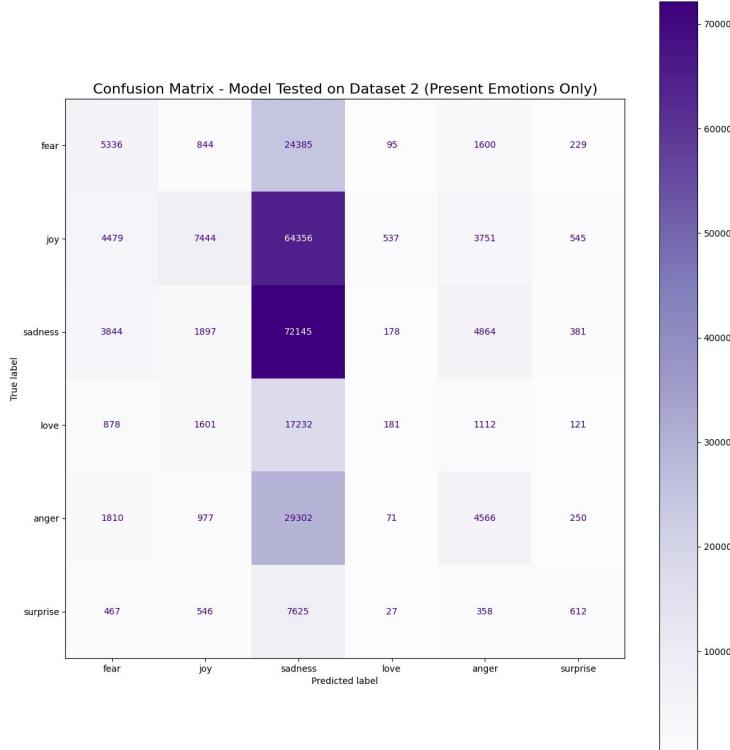
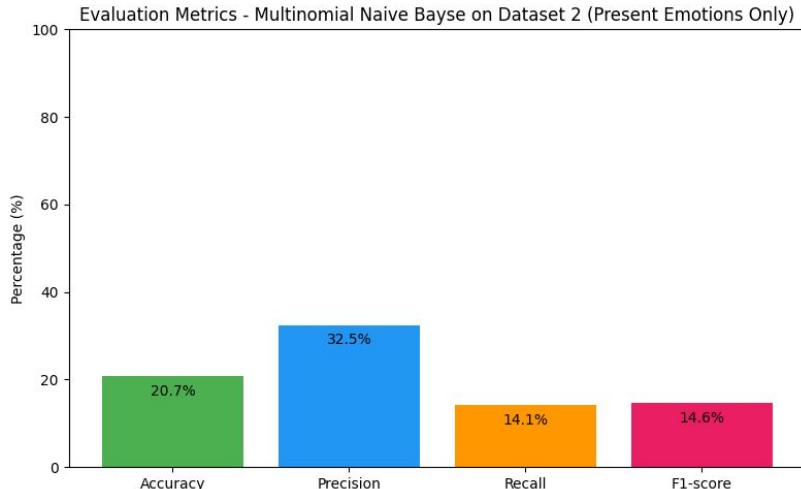
Performanță

- Acuratețe: **20.67%**
- Precizie: **32.43%**
- Recall: **14.09%**
- F1-Score: **14.53%**

Performanță pe clase de emoții

- Chiar și când evaluăm doar emoțiile prezente în Dataset 2, modelul arată **generalizare slabă**
- *Sadness* rămâne cea mai bine recunoscută emoție (**F1: 0.42**)
- *Joy* are cea mai mare precizie (**0.56**) dar recall foarte scăzut (**0.05**)
- *Fear și anger* arată performanță mai bună (**F1: 0.16 și 0.12**)
- *Love* performează extrem de slab (**F1: 0.01**)
- Modelul prezintă o **puternică tendință spre predicția sadness**, care domină rezultatele

Multinomial Naive Bayes



Multinomial Naive Bayes

Experiment 4: Model antrenat pe Dataset 2 și testat pe Dataset 2

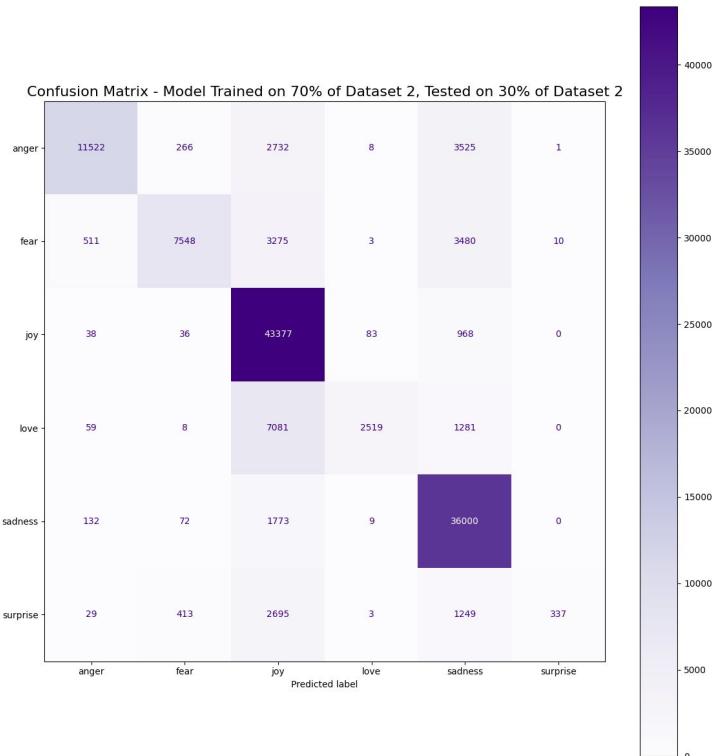
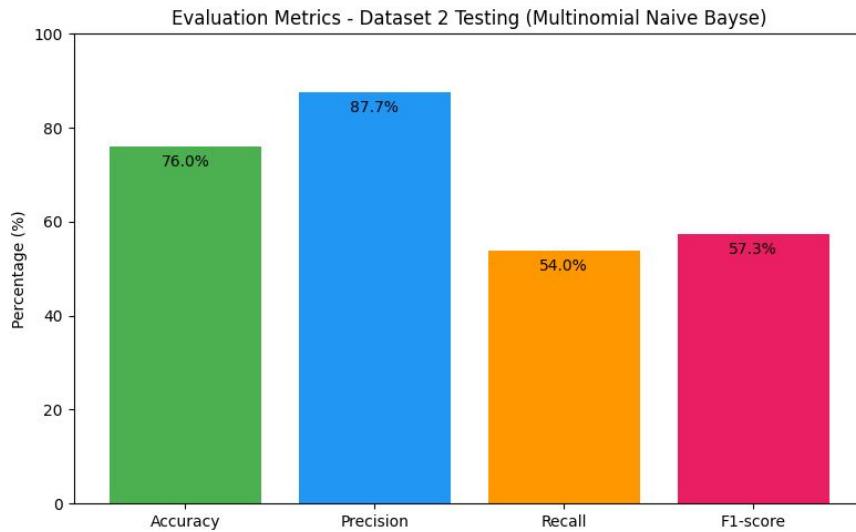
Performanță

- Acuratețe: **77.31%**
- Precizie: **87.60%**
- Recall: **56.19%**
- F1-Score: **59.83%**

Performanță pe clase de emoții

- **Cea mai bună performanță:**
 - Joy (**F1: 0.82**) - precizie de 0.71 și recall foarte ridicat (**0.97**)
 - Sadness (**F1: 0.85**) - precizie de 0.77 și recall ridicat (**0.95**)
- **Cea mai slabă performanță:**
 - Surprise (**F1: 0.13**) - precizie de 0.97 și recall foarte scăzut (**0.07**)
 - Love (**F1: 0.37**) - precizie de 0.96 și recall scăzut (**0.23**)

Multinomial Naive Bayes



Multinomial Naive Bayes

Experiment 5: Model antrenat pe Dataset 2 și testat pe Dataset 1 (toate emoțiile)

Performanță

- Acuratețe: **21.07%**
- Precizie: **47.48%**
- Recall: **36.49%**
- F1-Score: **21.55%**

Performanță pe clase de emoții

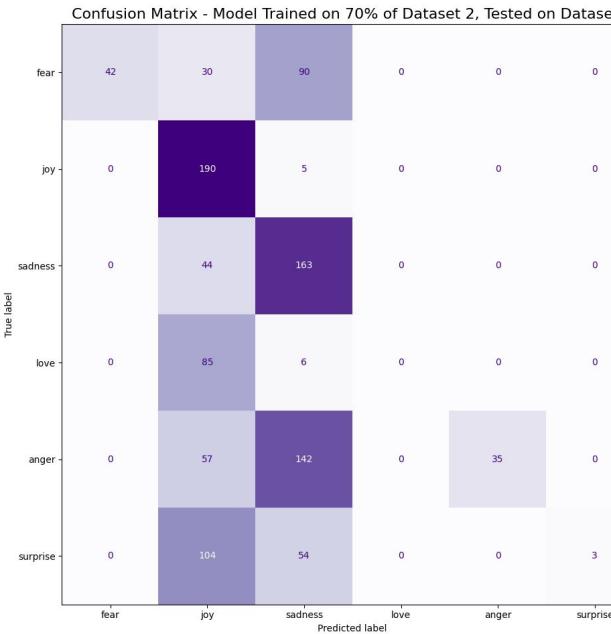
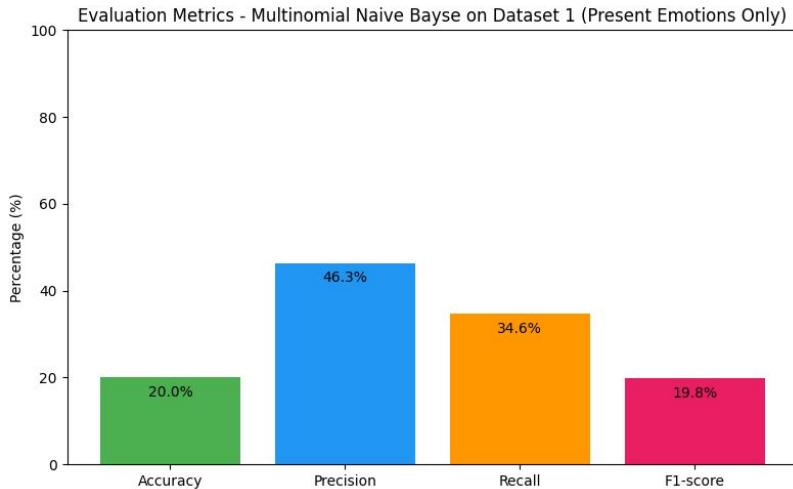
Cea mai bună performanță:

- **Fear (F1: 0.41)** - precizie excelentă (**0.98**) dar recall moderat (**0.26**)
- **Sadness (F1: 0.34)** - precizie moderată (**0.22**) dar recall ridicat (**0.79**)

Cea mai slabă performanță:

- **Love (F1: 0.00)** - modelul eșuează complet în recunoașterea acestei emoții
- **Surprise (F1: 0.04)** - deși are precizie perfectă (**1.00**), recall-ul este aproape inexistent (**0.02**)

Multinomial Naive Bayes



Random Forest Classifier

Descriere:

- Creează mai mulți **arbori de decizie** pe subseturi diferite ale datelor;
- Fiecare arbore votează → se alege clasa cu **vot majoritar**;
- Fiecare arbore învață dintr-un eșantion aleator + subset de caracteristici;
- Funcționează bine și pe date **zgomotoase sau dezechilibrate**.

Caracteristici cheie:

- Reduce overfitting-ul față de un singur arbore;
- Poate gestiona seturi de date **cu multe caracteristici**.

Random Forest Classifier

Experiment 1: Model antrenat pe Dataset 1 și testat pe Dataset 1

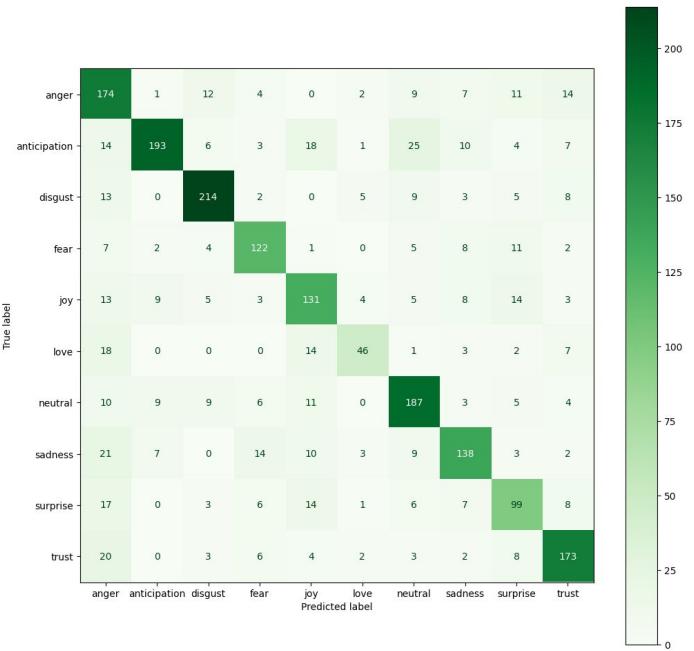
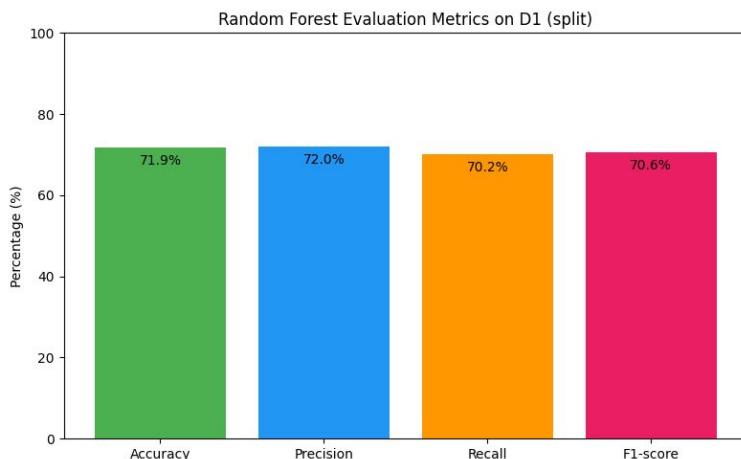
Performanță

- Acuratețe: **71.87%**
- Precizie: **71.97%**
- Recall: **70.17%**
- F1-Score: **70.63%**

Performanță pe clase de emoții

- **Cea mai bună performanță:** *disgust*
 - Precizie și recall foarte ridicate: **0.84/0.83**
 - Performanță constantă
- **Cea mai slabă performanță:** *love*
 - Recall foarte scăzut: **0.51**
 - Modelul ratează un număr semnificativ de exemple → cel mai slab scor F1: **0.59**

Random Forest Classifier



Random Forest Classifier

Experiment 2: Model antrenat pe Dataset 1 și testat pe Dataset 2 (toate emoțiile)

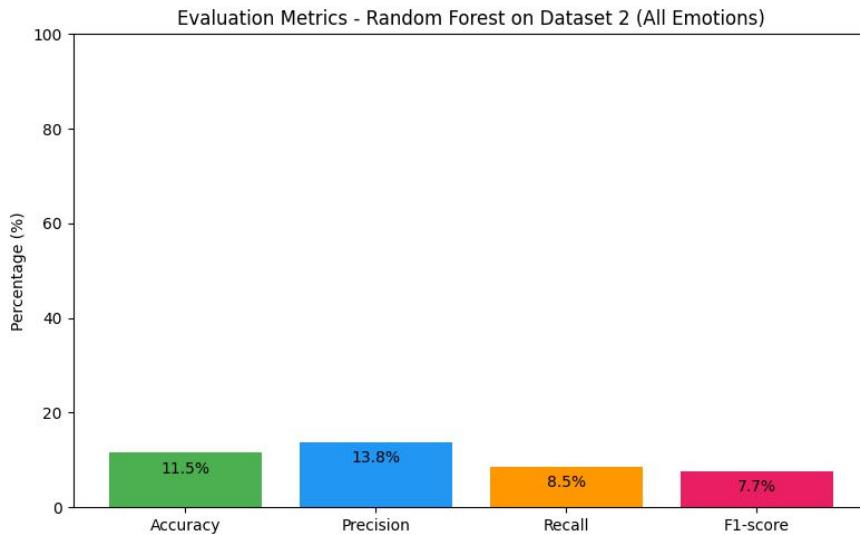
Performanță

- Acuratețe: **11.54%**
- Precizie: **13.75%**
- Recall: **8.51%**
- F1-Score: **7.68%**

Performanță pe clase de emoții

- Modelul are o performanță foarte slabă pe acest set, cu un scor F1 general de doar **11.16%**
- Emoțiile *anticipation*, *disgust*, *neutral* și *trust* nu sunt detectate deloc (**F1 = 0.00**)

Random Forest Classifier



Random Forest Classifier

Experiment 3: Model antrenat pe Dataset 1 și testat pe Dataset 2 (doar emoțiile comune dintre cele 2 dataset-uri)

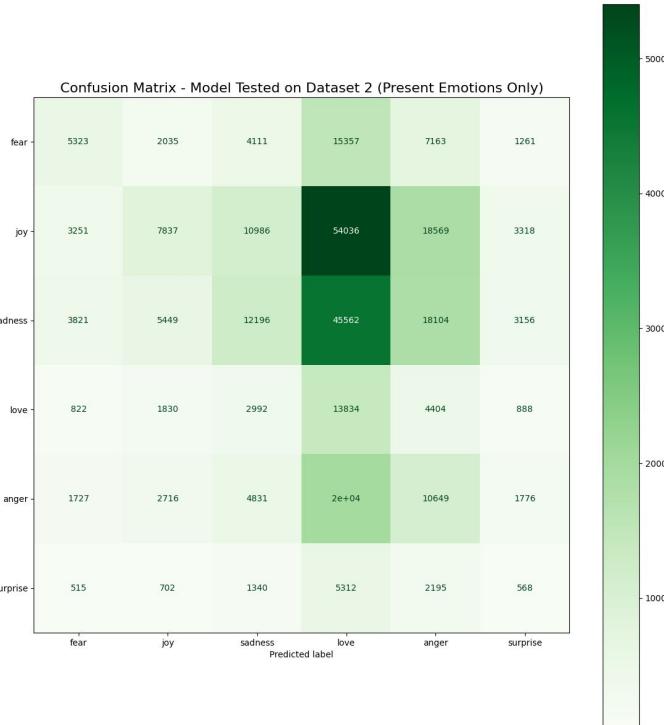
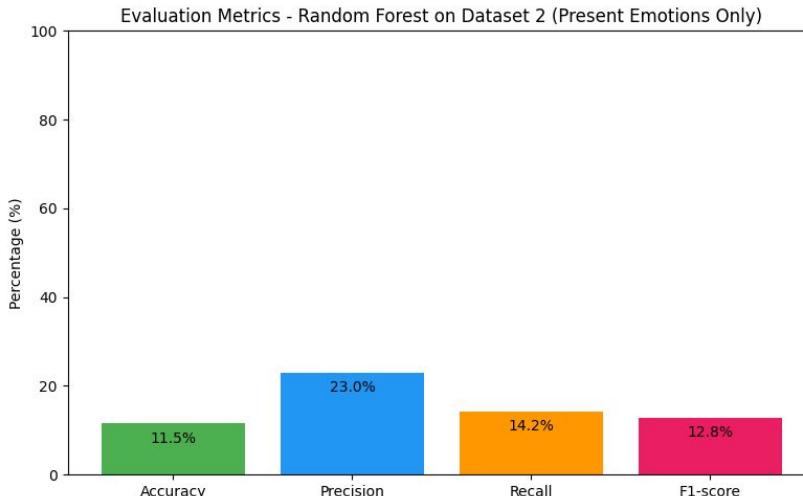
Performanță

- Acuratețe: **11.54%**
- Precizie: **22.9%**
- Recall: **14.18%**
- F1-Score: **12.81%**

Performanță pe clase de emoții

- Chiar și când evaluăm doar emoțiile prezente în Dataset 2, modelul are o **generalizare slabă**
- Cele mai mari scoruri F1: *anger* – **0.18**, *fear* – **0.16**
- Emoții frecvente precum *joy* (**0.09**) și *surprise* (**0.04**) sunt **clasificate foarte slab**
- Performanța generală rămâne **sub nivelul acceptabil**

Random Forest Classifier



Random Forest Classifier

Experiment 4: Model antrenat pe Dataset 2 și testat pe Dataset 2

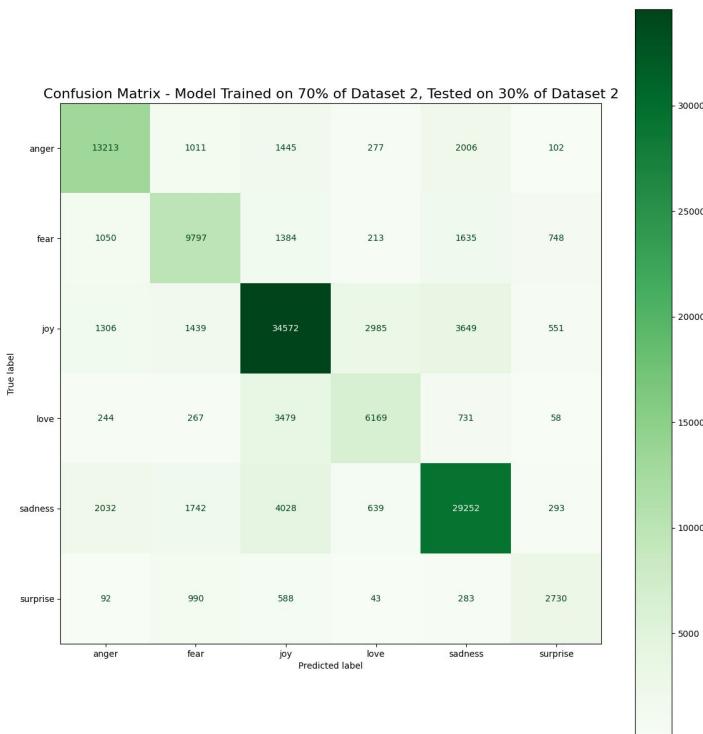
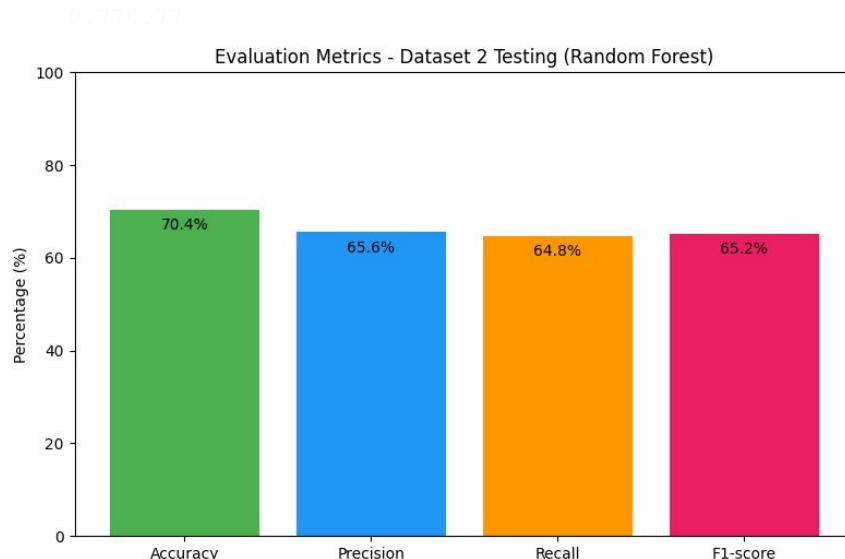
Performanță

- Acuratețe: **73.05%**
- Precizie: **68.74%**
- Recall: **68.01%**
- F1-Score: **68.25%**

Performanță pe clase de emoții

- **Cea mai bună performanță:** *joy*
 - Precizie și recall foarte ridicate: **0.78**
 - Performanță constantă și fiabilă
- **Cea mai slabă performanță:** *love*
 - Precizie și recall mai căzut: **0.56**
 - Modelul ratează un număr semnificativ de exemple → cel mai slab scor F1: **0.58**

Random Forest Classifier



Random Forest Classifier

Experiment 5: Model antrenat pe Dataset 2 și testat pe Dataset 1 (toate emoțiile)

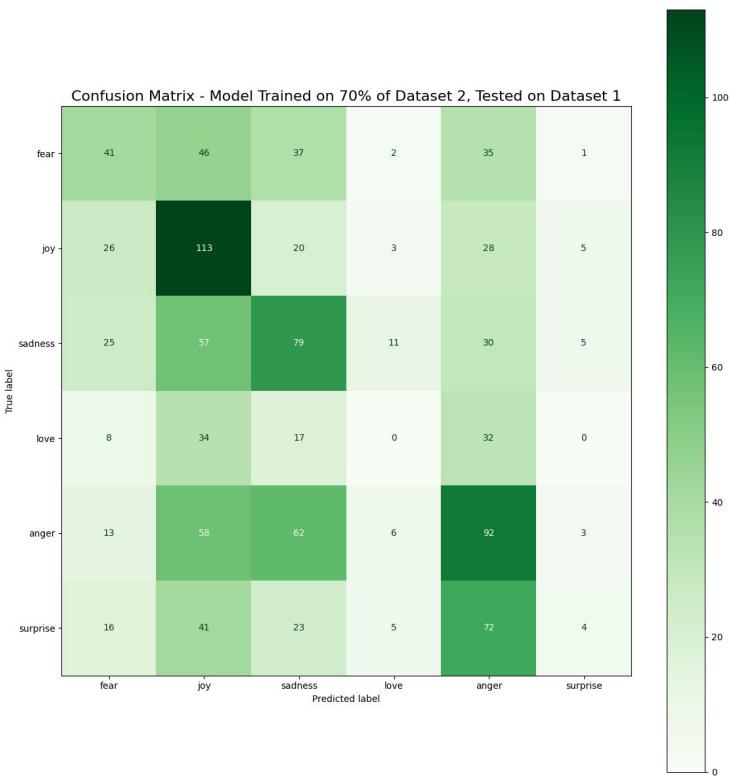
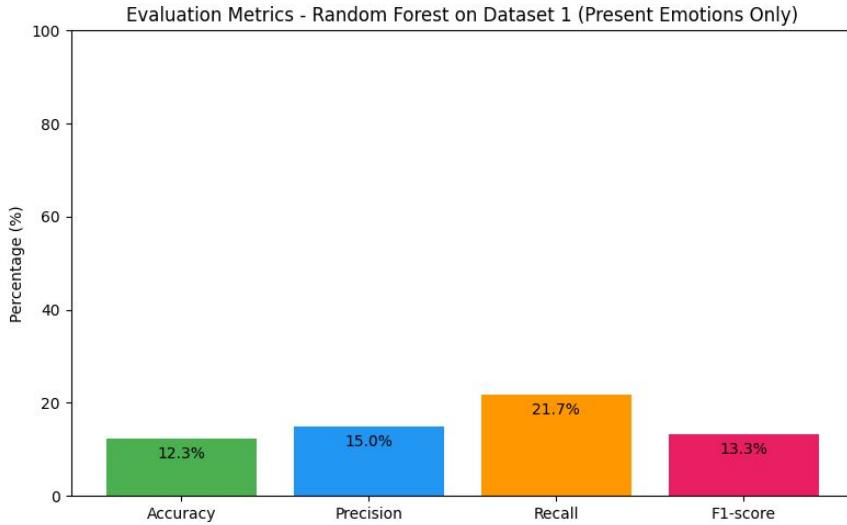
Performanță

- Acuratețe: **16.01%**
- Precizie: **14.00%**
- Recall: **27.20%**
- F1-Score: **16.14%**

Performanță pe clase de emoții

- *joy* are un recall ridicat (**0.58**), dar o precizie foarte scăzută (**0.14**) → mulți fals pozitivi
- *love* are cel mai slab clasificată (F1 0.00)

Random Forest Classifier



Logistic Regression

Descriere:

- Atribuie **ponderi (weights)** fiecărei caracteristici de intrare (ex: cuvintele din propoziție);
- Calculează un **scor pentru fiecare clasă** de emoții;
- Aplică funcția **softmax** pentru a transforma scorurile în **probabilități**;
- Alege clasa cu **cea mai mare probabilitate** ca predicție finală.

Caracteristici cheie:

- **Model probabilistic** – oferă scoruri de încredere pentru fiecare clasă;
- Ușor de implementat și **rapid de antrenat**;
- Funcționează bine când datele sunt **liniar separabile**.

Logistic Regression

Experiment 1: Model antrenat pe Dataset 1 și testat pe Dataset 1

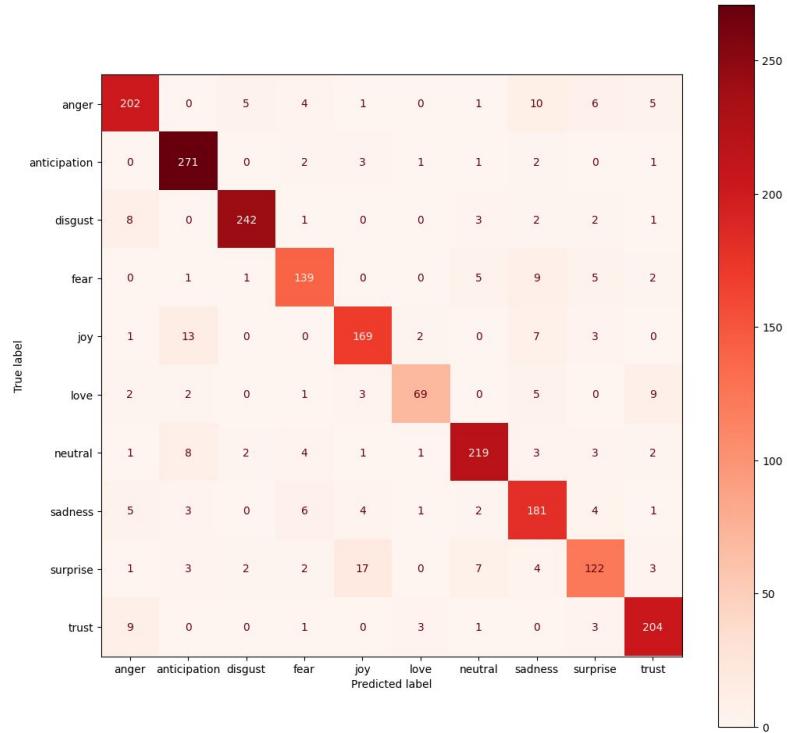
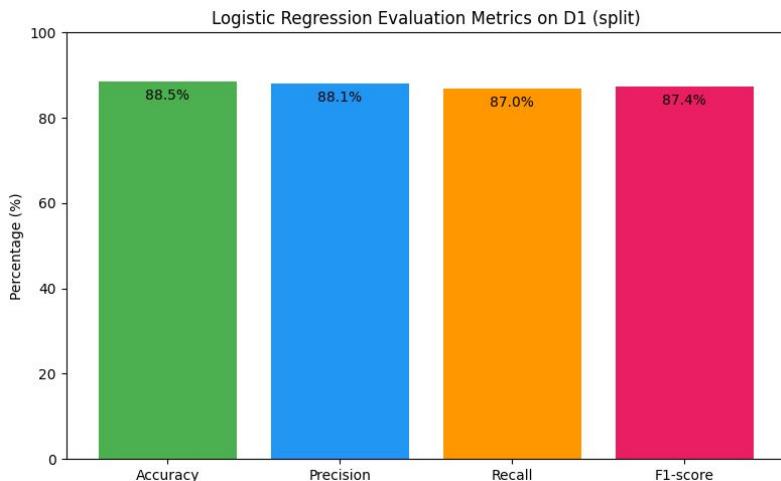
Performanță

- Acuratețe: **88.46%**
- Precizie: **88.08%**
- Recall: **86.97%**
- F1-Score: **87.42%**

Performanță pe clase de emoții

- **Cea mai bună performanță:** *disgust*
 - Precizia și recall foarte ridicate: **0.96/0.93**
 - Performanță constantă
- **Cea mai slabă performanță:** *surprise*
 - Deși precizia este acceptabilă, recall-ul scăzut (**0.76**) indică faptul că modelul ratează multe cazuri de “surprise”

Logistic Regression



Logistic Regression

Experiment 2: Model antrenat pe Dataset 1 și testat pe Dataset 2 (toate emoțiile)

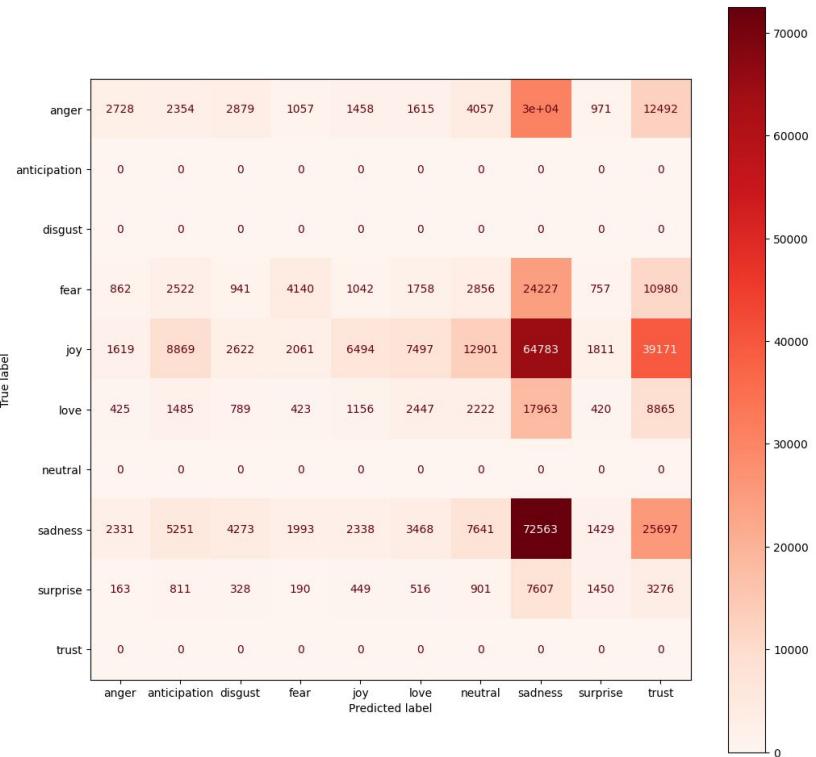
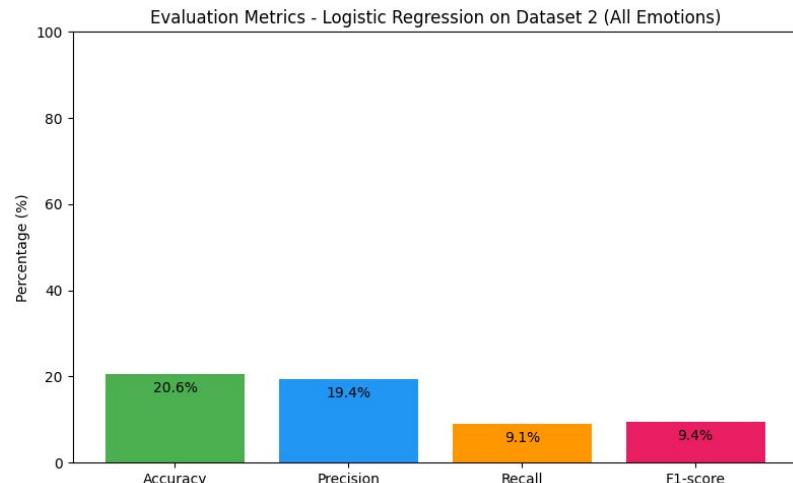
Performanță

- Acuratețe: **20.56%**
- Precizie: **19.44%**
- Recall: **9.03%**
- F1-Score: **9.40%**

Performanță pe clase de emoții

- Modelul are o performanță foarte slabă pe acest set, cu un scor F1 general de doar **9,40%**
- Emoțiile *anticipation*, *disgust*, *neutral* și *trust* nu sunt detectate deloc (**F1 = 0.00**)

Logistic Regression



Logistic Regression

Experiment 3: Model antrenat pe Dataset 1 și testat pe Dataset 2 (doar emoțiile comune dintre cele 2 dataset-uri)

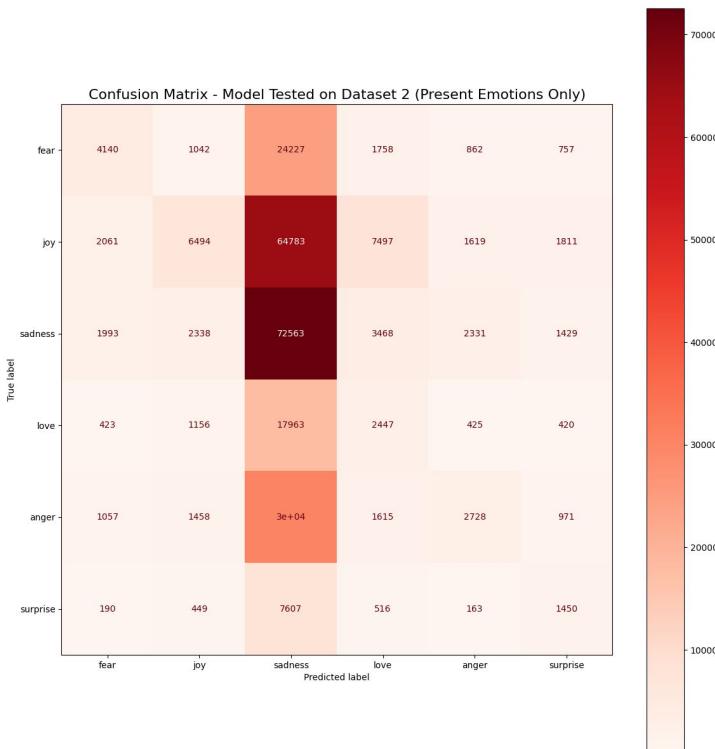
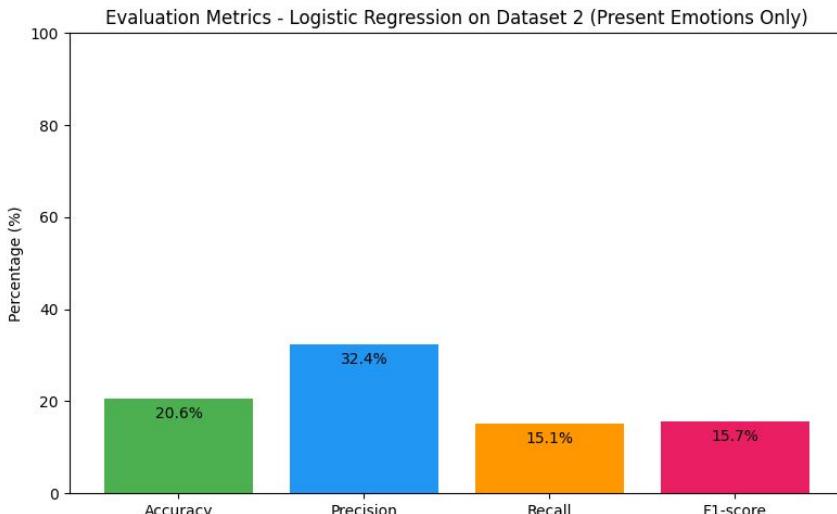
Performanță

- Acuratețe: **20.56%**
- Precizie: **32.40%**
- Recall: **15.05%**
- F1-Score: **15.64%**

Performanță pe clase de emoții

- Chiar și când evaluăm doar emoțiile prezente în Dataset 2, modelul are o **generalizare slabă**
- Cele mai mari scoruri F1: *sadness* – **0.42**, *fear* – **0.14**
- Emoții frecvente precum *joy* (**0.08**) și *anger* (**0.08**) sunt **clasificate foarte slab**
- Performanța generală rămâne **sub nivelul acceptabil**

Logistic Regression



Logistic Regression

Experiment 4: Model antrenat pe Dataset 2 și testat pe Dataset 2

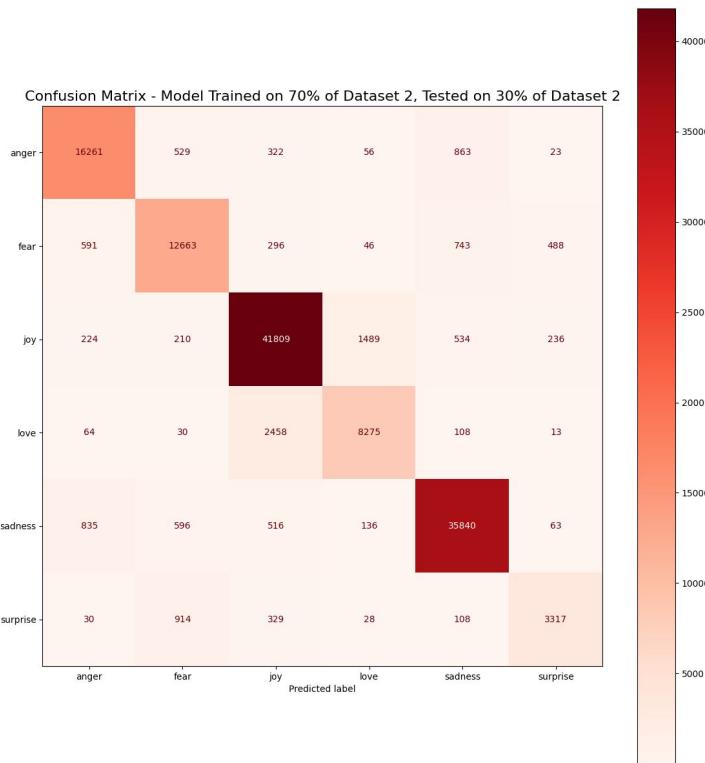
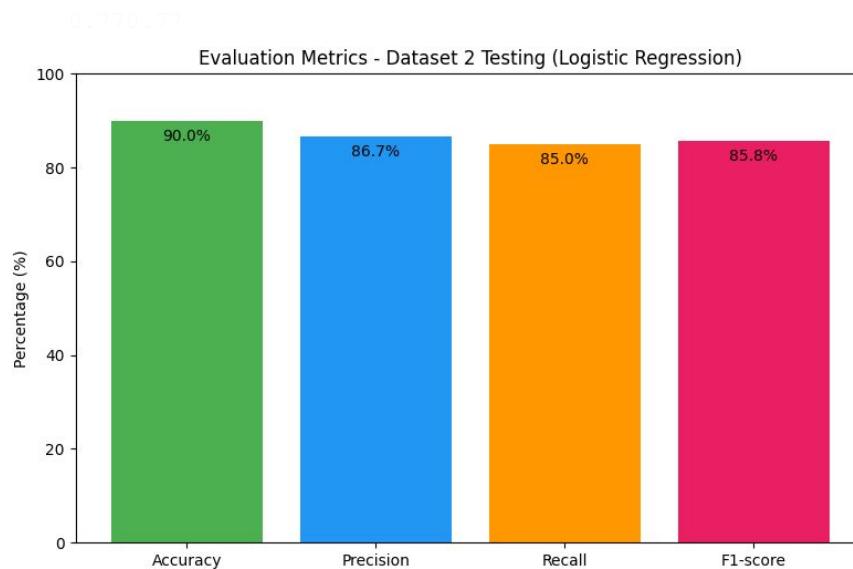
Performanță

- Acuratețe: **90.17%**
- Precizie: **87.15%**
- Recall: **84.92%**
- F1-Score: **85.95%**

Performanță pe clase de emoții

- Cea mai bună performanță: *sadness*
 - F1-score ridicat: **0.94**
 - Performanță constantă și fiabilă
- Cea mai slabă performanță: *surprise*
 - F1-score mai scăzut: **0.75**

Logistic Regression



Logistic Regression

Experiment 5: Model antrenat pe Dataset 2 și testat pe Dataset 1 (toate emoțiile)

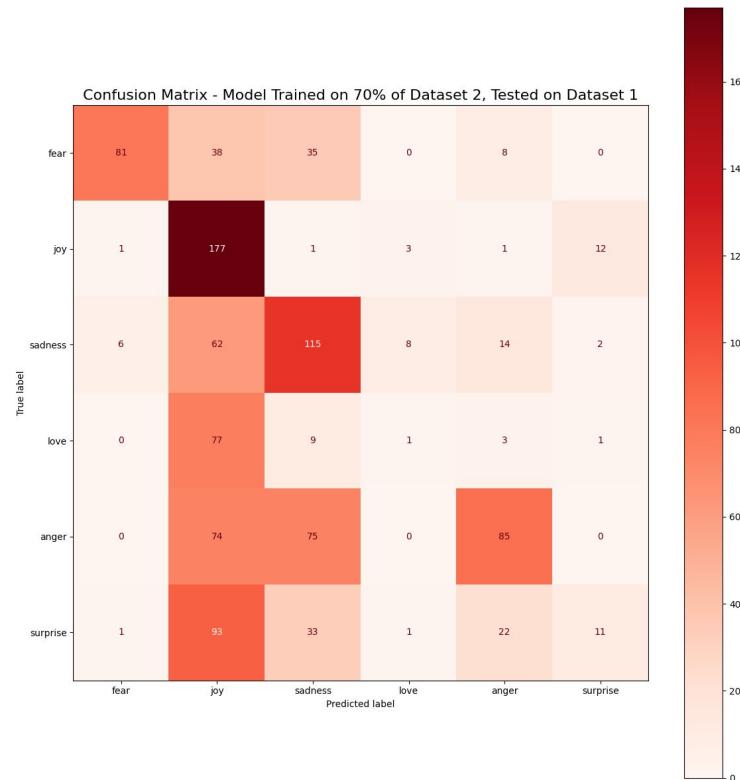
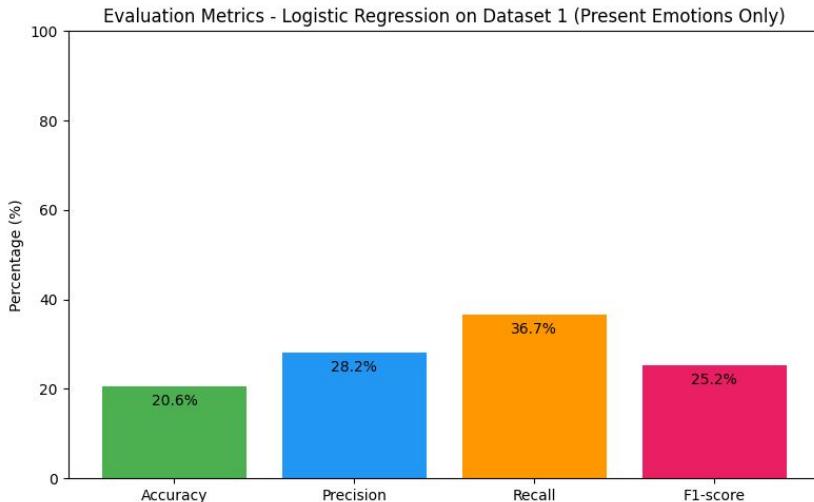
Performanță

- Acuratețe: **22.87%**
- Precizie: **31.00%**
- Recall: **40.09%**
- F1-Score: **27.89%**

Performanță pe clase de emoții

- *joy* are un **recall ridicat (0.91)**, dar o **precizie foarte scăzută (0.16)** → mulți fals pozitivi
- *love* are **cel mai slab clasificată (F1 0.02)**

Logistic Regression



Support Vector Machine (SVM)

Descriere:

- Caută **hiperplanul optim** care separă clasele de emoții
- Folosește **reprezentări vectoriale de dimensiuni mari** (ex. TF-IDF) pentru fiecare text
- Maximizează **distanța (margină)** între clase folosind *vectorii de suport*
- Eticheta este atribuită în funcție de **poziția vectorului** față de hiperplan

Caracteristici cheie:

- Se concentrează pe **separarea clară** a claselor
- Nu estimează probabilități – **model neprobabilistic**
- Robust la date cu dimensiuni mari și **bine definit pentru clasificare**

Support Vector Machine (SVM)

Experiment 1: Model antrenat pe Dataset 1 și testat pe Dataset 1

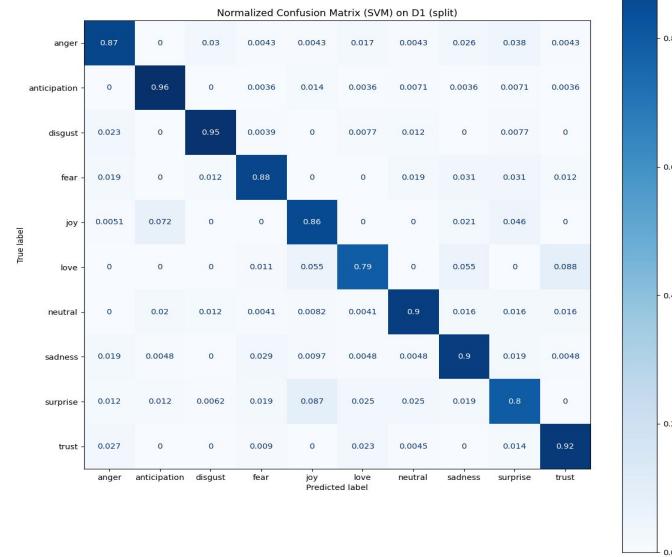
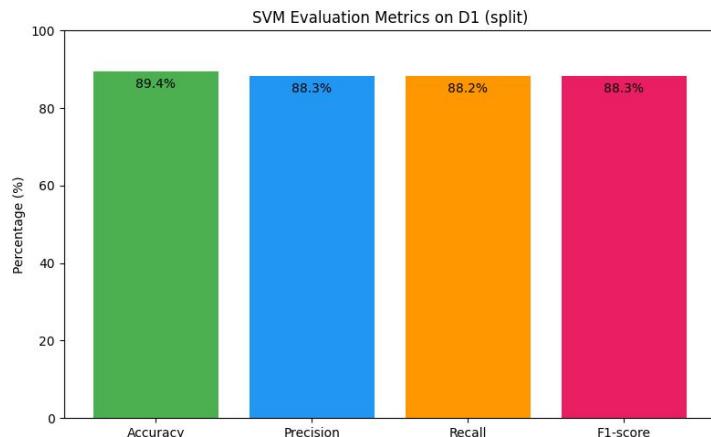
Performanță

- Acuratețe: **89.44%**
- Precizie: **88.31%**
- Recall: **88.22%**
- F1-Score: **88.25%**

Performanță pe clase de emoții

- **Cea mai bună performanță:** *disgust*
 - Precizie și recall foarte ridicate: **0.95**
 - Performanță constantă și fiabilă
- **Cea mai slabă performanță:** *surprise*
 - Precizie acceptabilă, dar recall scăzut: **0.77**
 - Modelul ratează un număr semnificativ de exemple → cel mai slab scor F1: **0.78**

Support Vector Machine (SVM)



Support Vector Machine (SVM)

Experiment 2: Model antrenat pe Dataset 1 și testat pe Dataset 2 (toate emoțiile)

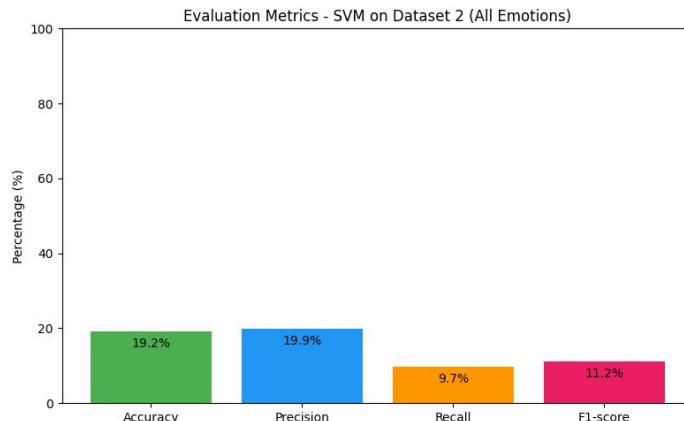
Performanță

- Acuratețe: **19.15%**
- Precizie: **19.93%**
- Recall: **9.71%**
- F1-Score: **11.16%**

Performanță pe clase de emoții

- Modelul are o performanță foarte slabă pe acest set, cu un scor F1 general de doar **11,16%**
- Emoțiile *anticipation*, *disgust*, *neutral* și *trust* nu sunt detectate deloc (**F1 = 0.00**)
- Acest rezultat indică o **incapacitate majoră de generalizare** la date noi

Support Vector Machine (SVM)



Support Vector Machine (SVM)

Experiment 3: Model antrenat pe Dataset 1 și testat pe Dataset 2 (doar emoțiile comune dintre cele 2 dataset-uri)

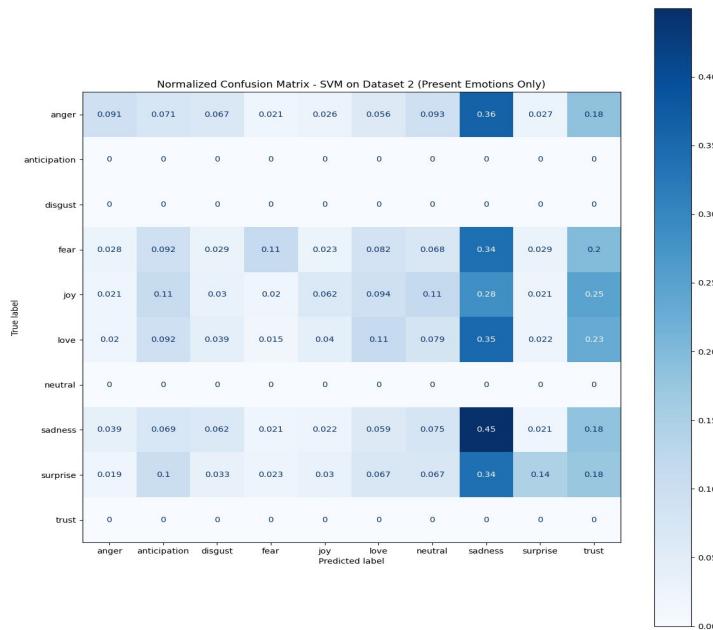
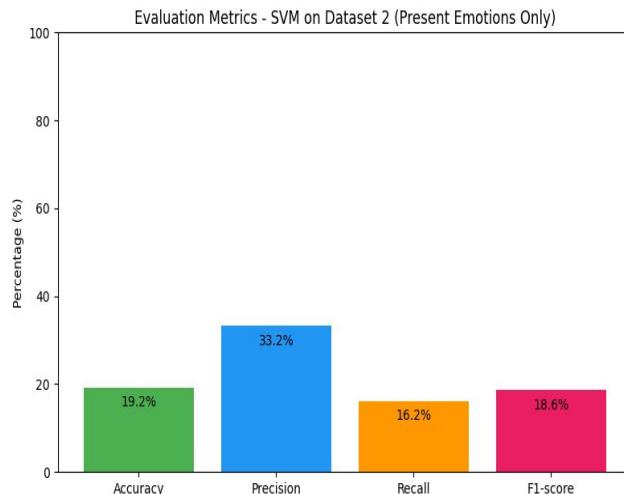
Performanță

- Acuratețe: **19.15%**
- Precizie: **33.21%**
- Recall: **16.19%**
- F1-Score: **18.60%**

Performanță pe clase de emoții

- Chiar și când evaluăm doar emoțiile prezente în Dataset 2, modelul are o **generalizare slabă**
- Cele mai mari scoruri F1: *sadness* – **0.40**, *fear* – **0.18**
- Emoții frecvente precum *joy* (**0.11**) și *anger* (**0.14**) sunt **clasificate foarte slab**
- Performanța generală rămâne **sub nivelul acceptabil**

Support Vector Machine (SVM)



Support Vector Machine (SVM)

Experiment 4: Model antrenat pe Dataset 2 și testat pe Dataset 2

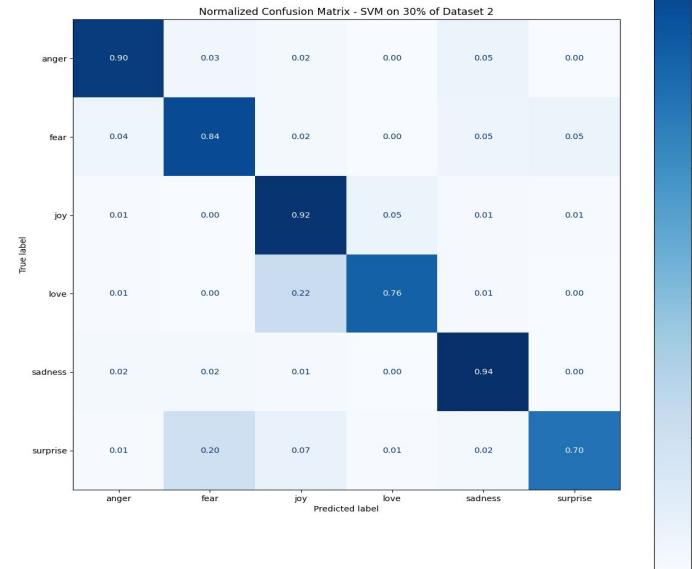
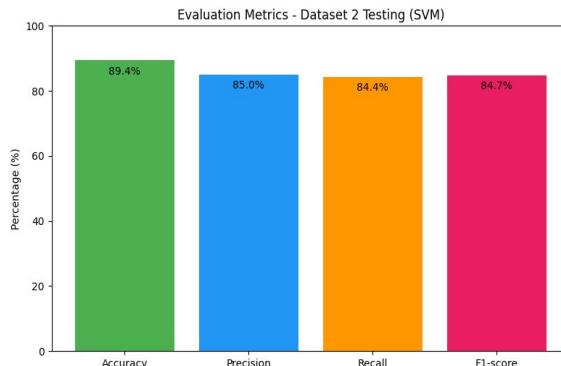
Performanță

- Acuratețe: **89.36%**
- Precizie: **85.04%**
- Recall: **84.39%**
- F1-Score: **84.70%**

Performanță pe clase de emoții

- **Cea mai bună performanță:** *sadness*
 - Precizie și recall foarte ridicate: **0.94**
 - Performanță constantă și fiabilă
- **Cea mai slabă performanță:** *surprise*
 - Precizie și recall mai căzut: **0.74**
 - Modelul ratează un număr semnificativ de exemple → cel mai slab scor F1: **0.72**

Support Vector Machine (SVM)



Support Vector Machine (SVM)

Experiment 5: Model antrenat pe Dataset 2 și testat pe Dataset 1 (toate emoțiile)

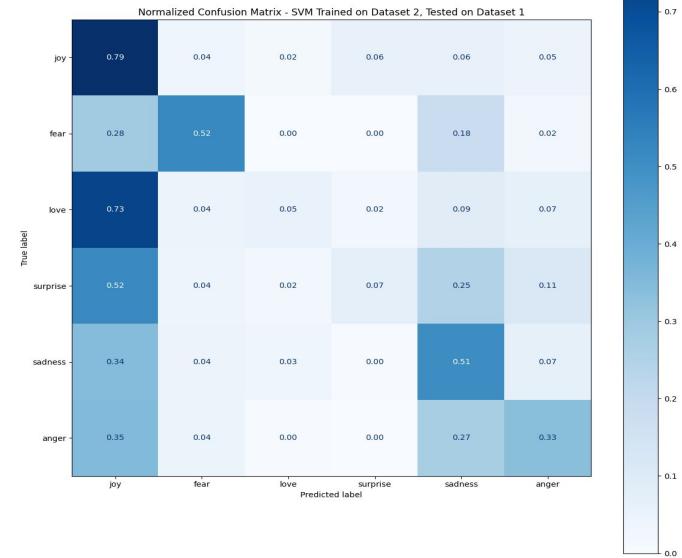
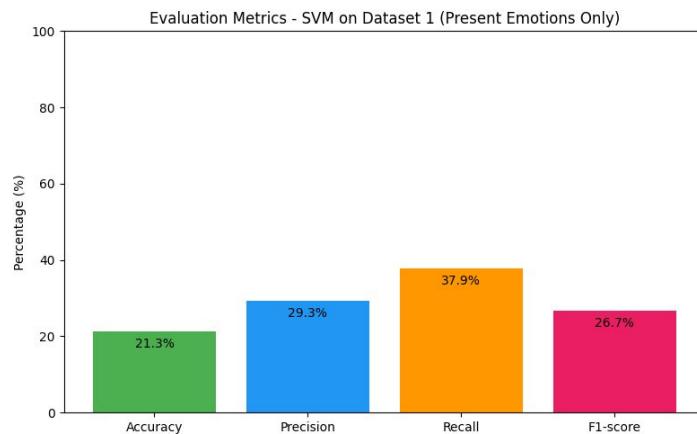
Performanță

- Acuratețe: **21.26%**
- Precizie: **29.28%**
- Recall: **37.86%**
- F1-Score: **26.74%**

Performanță pe clase de emoții

- *joy* are un **recall ridicat (0.79)**, dar o **precizie foarte scăzută (0.14)** → mulți fals pozitivi
- *fear* este cel mai echilibrat → F1-score de **0.55**
- *love* și *surprise* sunt **cel mai slab clasificate** (F1 sub 0.12)
- *sadness* și *anger* au scoruri moderate, dar inconsistent distribuite

Support Vector Machine (SVM)



BERT

BERT este un model avansat creat de Google care înțelege mai bine sensul cuvintelor, citind propozițiile în ambele direcții.

Cum funcționează:

- Citește textul **în ambele direcții** (stânga-dreapta și dreapta-stânga) pentru a înțelege mai bine contextul
- Împarte propozițiile **în tokeni**, apoi îi transformă **în ID-uri numerice**
- Cuvintele necunoscute sunt descompuse **în subcuvinte**

Token special:

- [CLS] este adăugat la începutul fiecărei propoziții
- După procesare, acesta conține **rezumatul propoziției** și este folosit pentru clasificare

Avantaje față de modelele clasice:

- Nu folosește trăsături selectate manual
- Învață direct din date
- Captează mai bine **nuanțele și sensurile subtile** din limbaj

BERT

Setup: Fine-tuning BERT pe Dataset 1

Configurație model:

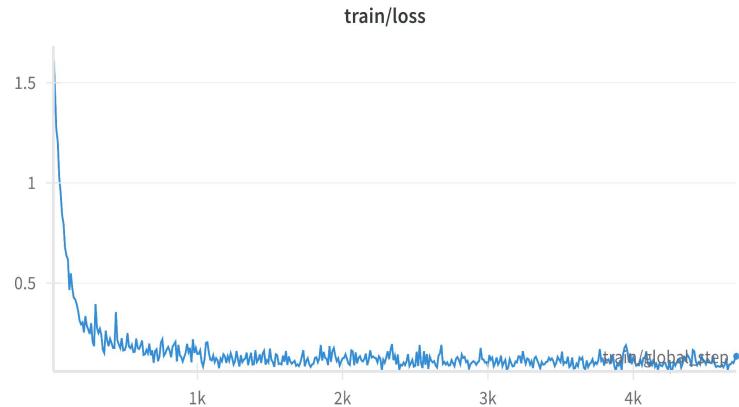
- Model folosit: bert-base-uncased (cu head de clasificare adaptat la numărul de emoții)
- Antrenare cu HuggingFace Trainer
- inițial: 3 epoci, batch size: 16, evaluare la fiecare 10 pași, dar pentru a reduce timpul de rulare am utilizat 1 epocă

Preprocesare date:

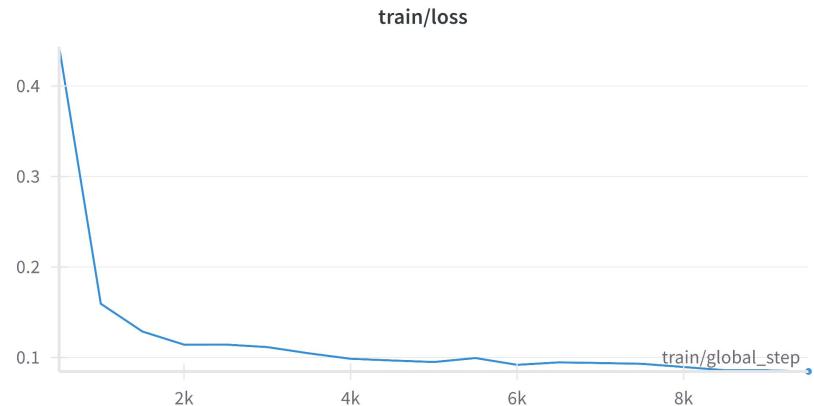
- Tokenizare cu lungime maximă de 128
- Padding și tăiere automată a secvențelor

BERT

BERT Dataset 1 - Training Loss Curve

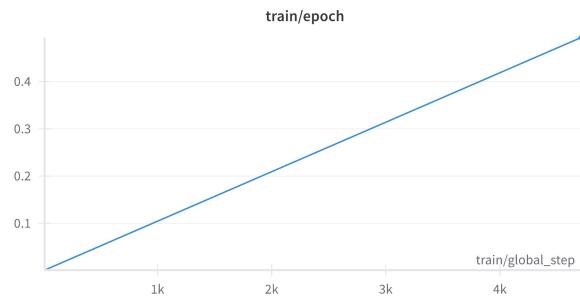


BERT Dataset 2 - Training Loss Curve

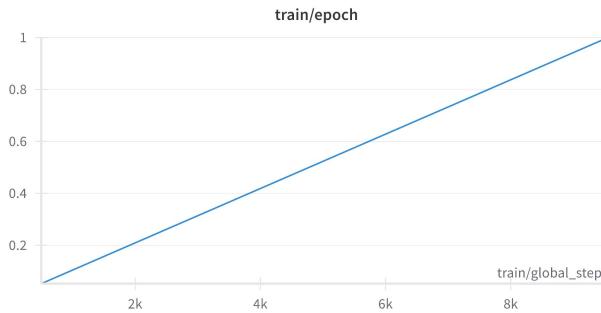


BERT

BERT Dataset 1 - Epoch Progression



BERT Dataset 2 - Epoch Progression



BERT

Experiment 1: Model antrenat pe Dataset 1 și testat pe Dataset 1

Performanță

- Acuratețe: **96.39%**
- Precizie: **96.41%**
- Recall: **96.11%**
- F1-Score: **96.23%**

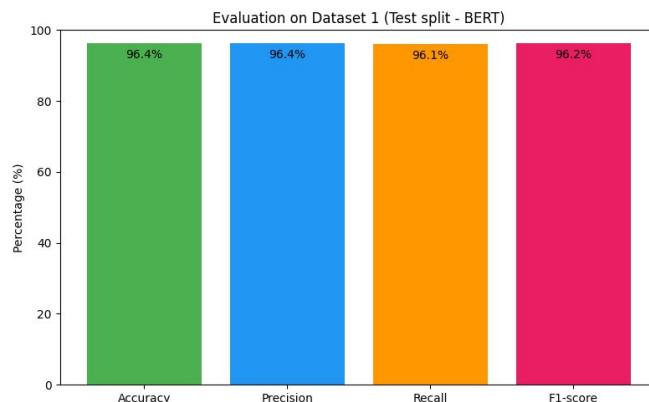
Performanță pe clase de emoții

Spre deosebire de modelele tradiționale (SVM, Logistic Regression), BERT oferă rezultate remarcabil de consistente, cu scoruri F1 de peste 0.90 pentru toate emoțiile.

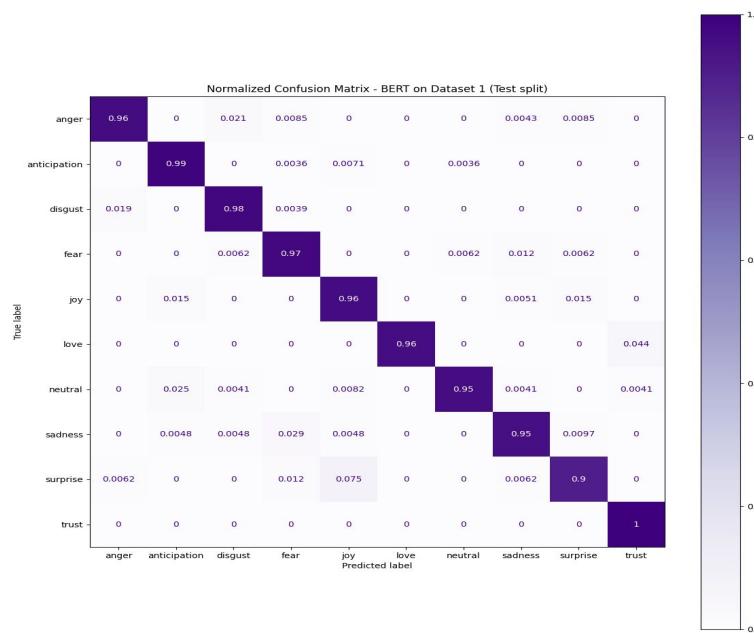
- **Cea mai bună performanță:** love
 - precizie perfectă (1.00) și recall ridicat (0.96), indicând că modelul clasifică această emoție fără erori.
- **Cea mai stabilă:** trust
 - Recall de 100% și precizie de 0.98 – un echilibru excelent între acuratețe și acoperire.
- **Cea mai slabă (dar tot puternică) :** surprise
 - F1-score de 0.92, afectat ușor de un recall mai scăzut (0.90), dar performanța rămâne foarte bună.

Acste rezultate evidențiază capacitatea superioară a modelului BERT de a înțelege subtilitățile limbajului emoțional, mult peste nivelul atins prin metode clasice.

BERT



Evaluation on Dataset 1 (Test split - BERT)



BERT

Experiment 2: Model antrenat pe Dataset 1 și testat pe un sample de 20k din Dataset 2 (doar emoțiile comune)

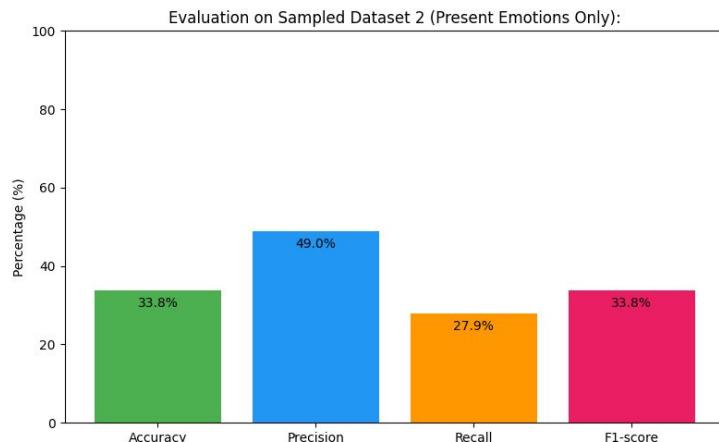
Performanță

- Acuratețe: **33.80%**
- Precizie: **48.95%**
- Recall: **27.90%**
- F1-Score: **33.84%**

Performanță pe clase de emoții

- Deși modelele tradiționale precum SVM și Logistic Regression au avut dificultăți majore (ex: SVM – F1-score ~11%), **BERT a reușit să dubleze performanța**, atingând un scor F1 macro de **33,84%**.
- Totuși, rezultatele arată că modelul **nu generalizează fiabil** pe Dataset 2.
- BERT oferă o **îmbunătățire clară** față de metodele clasice, dar **robustea sa cross-domain rămâne limitată**, mai ales în absența fine-tuning-ului pe date din domeniul țintă.

BERT



BERT

Experiment 3: Model antrenat pe Dataset 2 și testat pe Dataset 2

Performanță

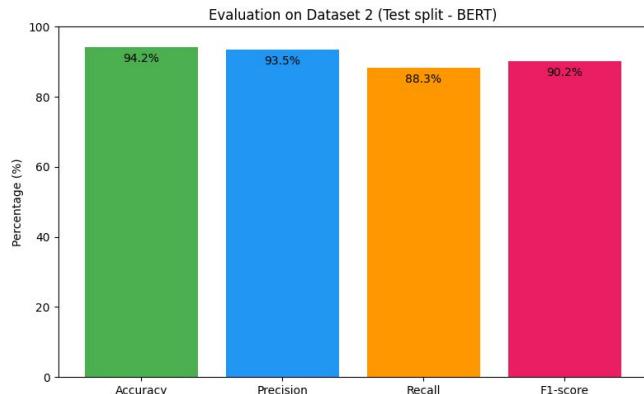
- Acuratețe: **94.15%**
- Precizie: **93.52%**
- Recall: **88.27%**
- F1-Score: **90.23%**

Performanță pe clase de emoții

BERT oferă performanță ridicată pe toate clasele de emoții, cu o acuratețe generală mare și un scor F1 macro de 90%.

- Joy și Sadness sunt clasificate cu **mare fiabilitate** (F1: 0.96–0.98)
- Fear are un **F1 de 0.91**, cu recall foarte bun
- Love ieșe în evidență prin **precizie perfectă (1.00)**, dar cu recall mai scăzut (ratează unele exemple)
- Surprise rămâne cea mai dificilă clasă, cu **cel mai mic F1: 0.80**

BERT



BERT

Experiment 4: Model antrenat pe Dataset 2 și testat pe Dataset 1 (doar emoțiile comune)

Performanță

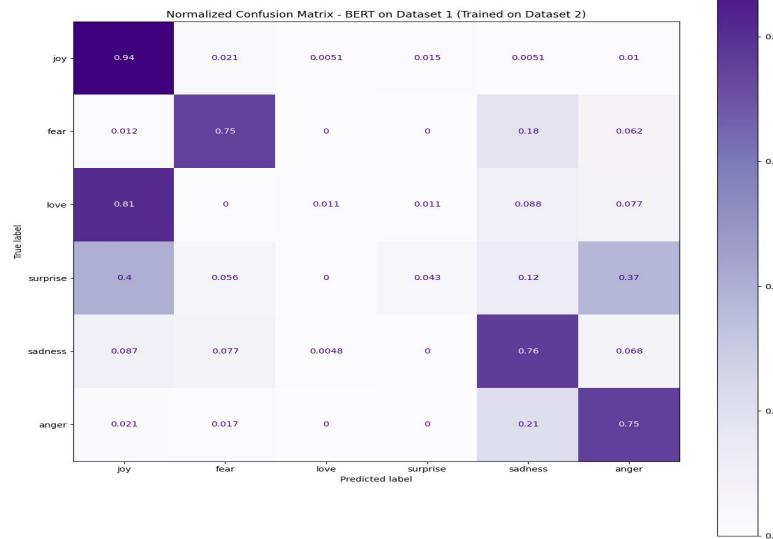
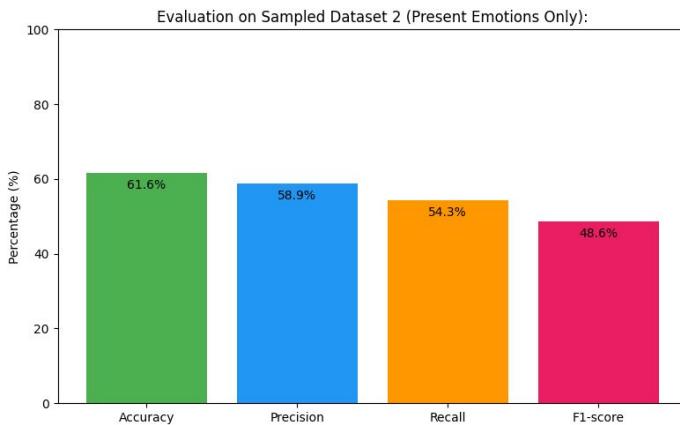
- Acuratețe: **61.61%**
- Precizie: **58.91%**
- Recall: **54.34%**
- F1-Score: **48.59%**

Performanță pe clase de emoții

Comparativ cu modelele tradiționale precum SVM (F1 ~26% în scenarii cross-dataset), **BERT demonstrează o generalizare semnificativ îmbunătățită.**

- Emoții recunoscute cu acuratețe ridicată:
 - *joy* – 94%
 - *sadness* – 76%
 - *anger* – 75%
 - *fear* – 75%
- Emoții mai problematice: *love* și *surprise* sunt frecvent confundate cu *joy* sau *anger*

BERT



BERT

Experiment 5: Modele antrenate pe Dataset 1 și Dataset 2 testate pe input-uri personalizate

Am testat ce rezultate se obțin cu ambele modele pentru următoarele texte:

- "I'm feeling overwhelmed and anxious today." – ambele prezic fear
- "Wow, I didn't expect that at all!" – ambele prezic joy
- "I can't believe how unfairly I was treated!" – ambele prezic anger
- "I care about you more than words can say." – ambele prezic joy (aici mai potrivit ar fi love)

Rezultate finale

Performanța modelelor în evaluarea in-domain (antrenare și testare pe același dataset)

Model	Dataset 1	Dataset 2
Multinomial Naive Bayes	83.13%	59.83%
Random Forest Classifier	70.63%	68.25%
Logistic Regression	87.42%	85.95%
Support Vector Machine	88.25%	84.70%
BERT	95.89%	91.74%

Rezultate finale

Performanță modelelor în evaluarea cross-domain (antrenare pe un dataset, testare pe celălalt)

Model	Dataset 1 → Dataset 2	Dataset 2 → Dataset 1
Multinomial Naive Bayes	8.7%	21.50%
Random Forest Classifier	7.68%	16.14%
Logistic Regression	9.40%	27.89%
Support Vector Machine	11.16%	26.74%
BERT	31.01%	48.22%

Concluzii Rezultate

Modele clasice: Performanță bună în domeniul antrenat, dar generalizare slabă între seturi de date

- SVM a fost cel mai performant model clasic pentru clasificarea emoțiilor
- Multinomial Naive Bayes oferă un echilibru bun între eficiență și performanță

Deep Learning: BERT arată performanță semnificativ superioară în ambele scenarii

- Performanță cross-domain de 2-4 ori mai bună față de modelele tradiționale
- Capacitate superioară de a interpreta contextul și subtilitățile emoționale din text

Limitări principale:

- Diferențele de vocabular și expresii între seturile de date
- Categorii de emoții diferite în Dataset 1 (10 emoții) vs Dataset 2 (6 emoții)
- Costuri computaționale ridicate pentru modelele transformer precum BERT
- Dezechilibrul de clase în ambele seturi de date

Limitări

Una dintre principalele limitări ale acestei lucrări este costul computațional asociat cu utilizarea modelelor de tip transformer, precum BERT, care necesită resurse semnificative de GPU pentru antrenare. Acest lucru poate limita scalabilitatea și accesibilitatea, în special pentru utilizatorii care nu dispun de hardware performant.

Modelele clasice de învățare automată, precum Support Vector Machines (SVM), Multinomial Naive Bayes, Random Forest și Logistic Regression pot fi antrenate eficient folosind doar CPU-uri.

Toate experimentele au fost realizate în Google Colab, care utilizează implicit CPU pentru antrenarea modelelor clasice. Totuși, pentru fine-tuning-ul modelului BERT, am schimbat mediul de rulare pentru a folosi GPU, deoarece modelele de tip transformer necesită o putere de calcul semnificativ mai mare.

Concluzii

Prin acest proiect, am învățat cum să aplicăm diverse algoritmi de învățare automată, atât modele clasice, cât și tehnici de tip deep learning precum BERT. Am dobândit experiență practică în preprocesarea textului, feature extraction și evaluarea modelelor folosind metrii precum acuratețea, precizia, recall-ul și scorul F1.

Ne-a plăcut în mod special să explorăm cum se comportă diferitele modele pe seturi de date diferite și să înțelegem provocările generalizării între domenii (cross-domain). Ce ni s-a părut cel mai interesant a fost modelul BERT și modul în care embedding-urile contextuale pot îmbunătăți performanța.

Pe de altă parte, am considerat frustrant faptul că fine-tuning-ul pentru BERT este destul de consumator de timp și necesită resurse hardware ridicate (precum GPU).