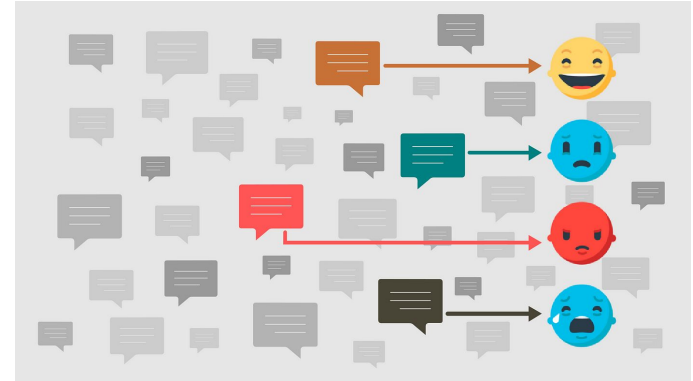


MULTI-LABEL EMOTION DETECTION IN TEXT

Irene Murua Txintxurreta
CISC 5800
5/12/2025

OBJECTIVE

- Build a model that detects up to 28 emotions in a single sentence.
- Compare classical TF-IDF + SVM with transformer fine-tuning (BERT, RoBERTa).
- Deliver a real-time web demo anyone can try.



DATA & APPROACH

- 58 k GoEmotions (Reddit, 28 labels)
- Pipelines: TF-IDF + SVM → BERT → RoBERTa
- Multi-label loss (BCE) + 0.5 threshold

Positive		Negative		Ambiguous
admiration 🙌	joy 😄	anger 😡	grief 😞	confusion 😕
amusement 😂	love ❤️	annoyance 😠	nervousness 😰	curiosity 🤔
approval 👍	optimism 🙌	disappointment 😞	remorse 😞	realization 💡
caring 🤗	pride 😊	disapproval 🗨️	sadness 😞	surprise 😲
desire 😍	relief 😌	disgust 🤢		
excitement 🤩		embarrassment 😳		
gratitude 🙏		fear 😨		

GoEmotions taxonomy: Includes 28 emotion categories, including "neutral".

RESULTS

Model	Precision	Recall	F1
TF-IDF + SVM	0.70	0.35	0.46
BERT	0.72	0.47	0.56
ROBERTA	0.73	0.47	0.56

High precision with moderate recall is typical—and acceptable—because the system prefers to be right when it does predict an emotion rather than list extra emotions that turn out to be wrong.

DOMAIN-SHIFT CHECK

Model	Top-1 acc	Top-3 acc
TF-IDF+SVM	0.09	0.21
BERT	0.19	0.39
ROBERTA	0.21	0.40

Top-1 checks if the model's first guess is correct; **Top-3** checks if any of its three best guesses are correct.

This test shows we need domain-adaptation techniques for real-world use.

CHALLENGES



- **Class imbalance** – Rare emotions like *grief* or *remorse* have very few training examples, so recall stays low.
- **Sarcasm & negation** – Phrases like “Yeah, perfect...” are often misread as positive.
- **Slang & emojis** – New words (“this slaps”) and emoji strings confuse the model, especially on Twitter.
- **Domain shift** – Style changes from Reddit to Twitter cut accuracy in half; we need adaptation tricks.

“These limits explain why recall is still moderate and why accuracy falls on tweets. They also guide our next steps: balanced sampling, sarcasm data, slang updates, and domain adaptation.”

LIVE DEMO

Emotion Classifier Demo

Paste in any text and see the top 3 predicted emotions (with scores).

text

These headphones sound incredible and I'm thrilled with the design, but I'm really disappointed that the Bluetooth keeps dropping.

Clear

Submit

output

admiration

admiration78%

disappointment66%

joy6%

Flag

KEY TAKEAWAYS

- **BERT and RoBERTa raise recall without sacrificing precision.**
The transformer models find many more true emotions yet stay just as accurate when they predict—an essential balance for reliable insights.
- **The performance drop on a different dataset highlights the need for domain-adaptation techniques.**
Accuracy falls when we move from Reddit comments to Twitter posts, so future work must teach the model to cope with new writing styles, hashtags, and emojis.
- **Next steps:**
 - Adapt the model to new text styles with extra domain-specific data
 - Expand training with rare emotions and sarcastic examples.
 - Explore slimmer, faster versions for easy deployment.



THANK YOU!

Irene Murua Txintxurreta
CISC 5800
5/12/2025

