Exposé

Working Title

"Politeness Detection in Online Communication: A Feature-Based Study of Text, Emojis, and GIFs"

Date: 01/09/2025

Adwaith Rajesh (1811915)

Surya Kiran (1807742)

Online communication lacks vocal and facial cues, making message tone easy to misinterpret. Users compensate with emojis and GIF-like visuals to signal friendliness, gratitude, sarcasm, or annoyance. While sentiment and sarcasm have been widely studied, politeness detection remains underexplored—especially in multimodal settings. This project builds a feature-based system that predicts **Polite / Neutral / Impolite** from text plus non-textual cues. We design an interpretable representation: **sentence embeddings for text**, **linguistically motivated features for emojis**, and **controlled semantic tags for visual reactions**. Although "GIFs" appear in the title to emphasize scope, we operationalize this visual dimension with stickers (common on WhatsApp/Telegram) for feasibility.

We evaluate with ablations and cross-validation and discuss applications in customer service, moderation, and workplace communication.

Motivation & Problem Statement

Politeness supports smooth interaction, conflict avoidance, and customer satisfaction—but is hard to detect from text alone. Emojis and visual reactions (GIFs/stickers) frequently shift perceived tone: 🙂 can soften a blunt request; 🙄 can turn neutral text impolite.

Motivation: Whenever I type messages on platforms like Telegram or WhatsApp, I always try to include emojis or stickers along with them to make my message more polite

In case of email, I always start my email with an emoji. I definitely feel like it always adds an extra in pleasing the receiver or helps in softening my message.

The use Stickers and Gif can be heavily seen in reddit messages and twitter tweets showing Softening disagreement / criticism or Showing empathy / support

Problem: Can we detect politeness reliably using **text + emoji + sticker signals**, and quantify each modality's contribution?

Research Objectives and Questions

- Incorporate multimodal signals — text, emojis, and GIF-like visuals — into a feature-based classification model.
- Compare the performance of text-only vs multimodal models.
- Identify which emojis or visual markers consistently soften or intensify tone.

1. Do emojis and stickers measurably improve politeness detection beyond text-only baselines?

2. Which cues most strongly affect predictions? (Text, Emoji, or Sticker Tags or Any combination)

3. Where do models fail (e.g., sarcasm without emojis, ambiguous stickers), and why?

Build a dataset of online messages annotated for politeness levels (polite, neutral, impolite).

Representation

Text: Sentence-BERT

Emojis: emoji2vec (Embeddings)

Eg: Give me the report now 🙂" → The 🙂 here is not "happy face," but a *softener* of an otherwise rude command.

Stickers: one-hot over the 8 controlled categories (interpretable).

Eg: categories = ["HAPPY", "SAD", "ANGRY", "SUPPORTIVE", "CELEBRATORY", "SARCASTIC", "CONFUSED", "NEUTRAL"]

Extra Cues:

Softeners 🙂😊🙏 … (make a message sound nicer)

Hostility markers 😡😒👎 … (make a message sound rude/impolite)

Intensifiers 😂🔥🎉 … (make emotions stronger)

Final vector: [text_emb| emoji vectors| sticker_onehot(8)].

Models

Primary: Logistic Regression

Logistic Regression concat everything:

Text Embedding (SBERT)	Emoji Embedding (emoji2vec)	Sticker One-Hot (8-dim)	Emoji Flags (3 features)	Label
`[0.21, -0.14, 0.33…]`	`[0.12, -0.45, 0.67…]`	`[0,0,1,0,0,0,0,0]`	`[soft=1, host=0, int=0]`	Polite

Ablations: Text-only; Text+Emoji; Text+ Text+Emoji+Sticker.

Dataset

id	text	emojis	sticker tags	label
1	Send me the file 🙂	🙂	ANGRY	impolite
2	Thanks a lot! 🎉	🎉	HAPPY	polite
3	That's fine 😒	😒	SARCASTIC	impolite
4	Appreciate your help 🙏	🙏	SUPPORTIVE	polite
5	You forgot again 😑	😑	CONFUSED	impolite

Expected Contributions

- A politeness-focused multimodal dataset (text+emoji+sticker tags).
- An interpretable baseline showing how non-text cues shift politeness perception.
- Empirical ablation evidence for the addition of emojis/stickers.

Applications

- Customer service (WhatsApp Business): detect annoyed/impolite replies and adapt tone.
- Online moderation: catch hostility masked by "polite" text but hostile emojis/stickers.
- Workplace chat (Slack/Teams): surface passive-aggressive exchanges.
- Education: help language learners understand pragmatic tone shifts.

Sample Expected Outcome

Input Used	Example Message	Model Output	Explanation for Result
Text only	"Send me the file 🙂"	Impolite ❌	Text looks like a command → model marks impolite.
Text + Emoji	"Send me the file 🙂"	Polite ✅	Emoji2vec captures 🙂 as "smile" → tone seems polite.
Text + Emoji + Sticker	"Send me the file 🙂 [SUPPORTIVE sticker]"	Polite ✅	Sticker category = SUPPORTIVE; Emoji flag = softener → both push prediction to polite.