When Emojis Speak Louder Than Words:

Multimodal Sarcasm Detection

Vasiliki Pantelopoulou & Panagiota Nalmpanti

ARISTOTLE UNIVERSITY OF THESSALONIKI

SCHOOL OF INFORMATICS

Data & Web Science MSc Program

Introduction

Sarcasm detection is difficult because the intended meaning often differs from the literal text. In this project, we build a text-only baseline inspired by Khan et al. using RoBERTa, attention and convolution blocks, and we compare it with a multimodal variant that also uses emoji-based visual features extracted with CLIP.

Both models are evaluated on iSarcasmEval to test whether emojis improve sarcasm detection.

Methodology

Data Preprocessing

- Dataset: SarcasmEval (English tweets)
- Split: Train and Validation (80/20), official Test set

We apply light text cleaning to keep transformer performance stable:

- URL removal
- whitespace normalization
- slang expansion
- raw tweet text for emoji extraction

Baseline Model

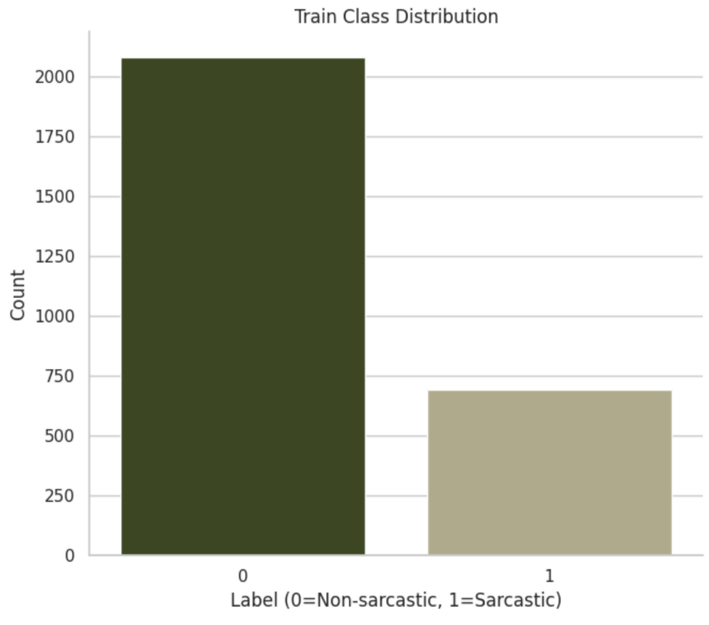
The baseline model follows a transformer-based text-only architecture.

- RoBERTa-base as the text encoder
- Multi-Head Attention blocks
- Depthwise Convolution blocks to capture local patterns
- Prediction using [CLS] token representation for classification

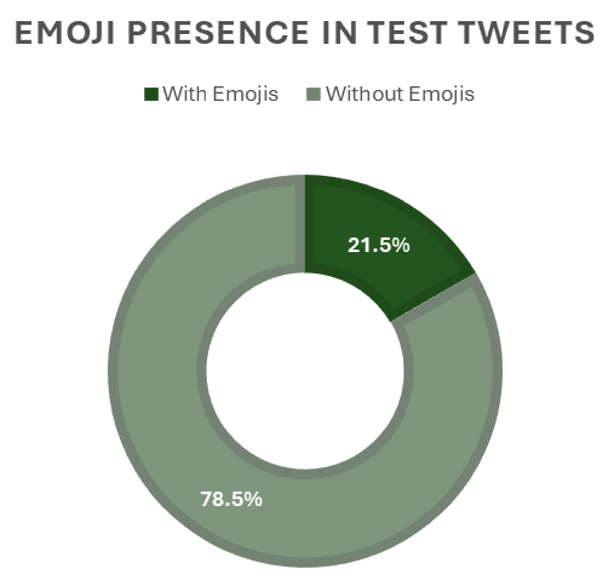
Proposed Variant

Extend the baseline with a simple multimodal variant that combines textual information with emoji-based visual features.

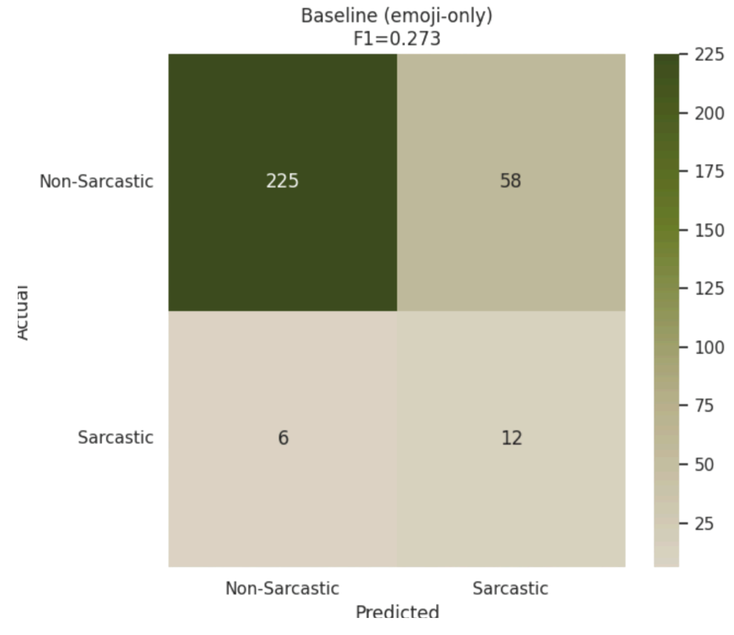
- Add an emoji-visual branch parallel to the text encoder
- Render emojis into a 224×224 image grid
- Extract visual features using a pretrained CLIP Vision Encoder



Train set is imbalanced, therefore F1-score is used as the main metric.

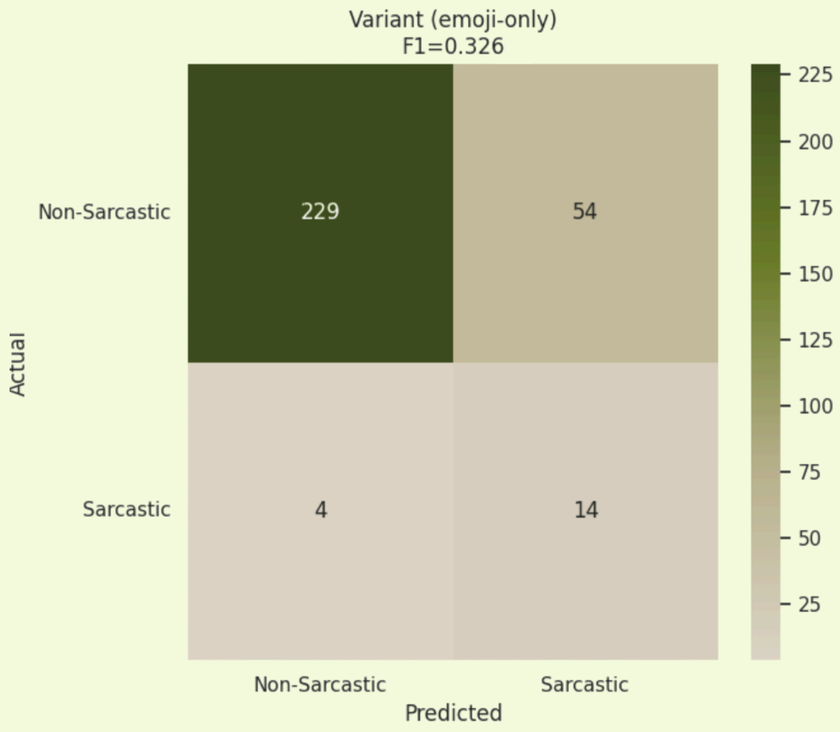


Top emojis: 😭 😂 ❤️ 🙃



Baseline achieves low sarcasm F1, indicating limited ability to exploit emoji-related cues.

Results



The multimodal variant achieves a higher sarcasm F1 (0.326), suggesting that emoji-based information improves predictions.

Golden Examples

1) "I just can't wait to spend time with my family over Christmas! I just love being the only single one and the many many questions asking when will I get a boyfriend 🙄"

2) "Top 10 pools in my book: 1. Swimming pool 2. Paddling pool 3. Above-ground pool 4. Family pool 5. Architectural pool 6. Indoor pool 7. Lap pool 8. Olympic size pool 9. Natural pool 10. Salt water pool Sorry Liverpool you are not top 10 pools in my book 😭😭😭"

True=1 Baseline=0 Variant=1

References

[1] Shumaila Khan, Iqbal Qasim, Wahab Khan, Khursheed Aurangzeb, Javed Ali Khan, and Muhammad Shahid Anwar. 2025. A novel transformer attention-based approach for sarcasm detection. Expert Systems 42, 1 (2025), e13686.