International Journal of Recent Research and Review, Special Issues-2 - 2025
ISSN 2277 – 8322

SAY WHAT YOU DON'T MEAN: SARCASM AND EMOTION ANALYSIS IN CODE-MIXED NLP

Vimal Daga
CTO, LW India |
Founder, #13 Informatics Pvt Ltd
LINUX WORLD PVT. LTD.

Preeti Daga
CSO, LW India |
Founder, LWJazbaa Pvt Ltd
LINUX WORLD PVT. LTD.

Shagun Saini
Research Scholar
LINUX WORLD PVT. LTD.

Abstract- Aren't people funny, typically confused about their own feelings, creating devices to understand sentiment and feeling in words? Such irony is at the core of sentiment analysis and affective computing. Even though emotion is a fundamental part of communication, as it gets more intricate when sarcasm, code-mixed text, and cultural context come into play. Human language is not often literal feelings may be concealed behind sarcasm, humor, or multilinguality, so sentiment identification is a naturally complicated task. Sentiment and emotion in language have real-world application in social media monitoring, customer service, political language, and psychiatric assessment. Sarcasm detection prevents mislabeling in opinion mining. Emotion recognition makes empathetic AI systems possible. Nonetheless, current models do not perform well in real-world, noisy settings where individuals blend languages, emotions are subtle, and sarcasm is rampant.Traditional sentiment analysis models depend on lexicon-based or supervised learning approaches, primarily trained on English data. Emotion detection has advanced with the employment of affective lexicons and deep learning. However, little work discusses the confluence of emotion, sarcasm, and multilingualism. Code-mixed corpora such as FIRE and LinCE provide benchmarks, but end-to-end solutions are not common.

Keywords: Sarcasm Detection, Emotion Classification, Bilingual NLP, Code-Mixed Language, Sentiment Analysis, Affective Computing, Contextual NLP

I. INTRODUCTION

Isn't it ironic that human beings, most of the time confused about their own feelings, are designing machines to recognize and interpret those very emotions? This irony is the essence of sentiment analysis and affective computing. Human emotions are rich, complex, and most of the time ambiguous. They can be conveyed through sarcasm, hidden behind humor, or buried in multilingual or code-mixed speech, rendering their identification a complex task even for humans, not to mention machines. As language gets increasingly casual, mixed, and context-dependent—especially on Twitter, YouTube, and WhatsApp—the requirement for machines to interpret emotion, sarcasm, and bilingual expression becomes much higher. The practical use of such interpretation in the real world is immense: from the analysis of political rhetoric and customer complaints to identifying indicators of mental illness.

Even with notable breakthroughs in Natural Language Processing (NLP), present sentiment analysis systems continue to falter in such noisy real-world settings. Most models are built upon monolingual, formal datasets and do not understand the multifaceted nature of emotional communication. Sarcasm tends to cause sentiment misclassification, and code-mixing language (such as Hinglish) has linguistic and structural issues that disrupt traditional NLP pipelines. This work bridges these gaps by investigating the nexus of emotion, sarcasm, and bilingualism in sentiment analysis. We seek to understand how contemporary NLP methods—specifically transformer-based models—can be modified to learn code-mixed emotional information with better nuance and cultural sensitivity. In the process, we not only improve machine comprehension but also consider the deeper irony of instructing artificial mechanisms to "feel" what we ourselves struggle to articulate.

II. LITERATURE REVIEW

The area of sentiment and emotional analysis has seen significant development, moving from simplistic polarity classification to more sophisticated emotion and rhetorical device identification such as sarcasm. Sentiment analysis techniques previously were mostly based on lexicon-based or rule-based techniques, including SentiWordNet and AFINN, that offered building block sentiment polarity identification. But these techniques were not able to identify contextual complexities. As machine learning, particularly supervised classifiers such as Naïve Bayes and Support Vector Machines (SVM), followed by deep models including Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs) emerged, the ability to learn patterns of sentiment grew immensely. In emotion detection, researchers moved beyond sentiment polarity to identify intricate emotional states with the aid of classification frameworks based on Ekman's emotions or the Plutchik wheel. Lexical resources such as Mohammad and Turney's NRC Emotion Lexicon were significantly able to facilitate the large-scale annotation of emotions. These advancements, however, were mostly concentrated on monolingual and grammatically formatted English text and hence were restricted in their applicability to actual environments where language is informal, code-mixed, and contextually stacked. Detection of sarcasm adds a further layer of difficulty. Being a linguistic phenomenon, sarcasm is typically characterized by some contrast between literal sentiment and intended meaning, which makes it challenging for models to recognize without contextual anchoring. Early attempts employed surface features like sentiment polarity reversals or overstatement. Sarcasm identification has been supported by corpora such as SARC (Reddit-based) and Twitter datasets, where noisy supervision comes in the form of hashtags (#sarcasm). Researchers such as Riloff et al. suggested semi-supervised approaches identifying sarcasm based on patterns of sentiment contrast, whereas others, including Ghosh et al., utilized deep neural models that tap conversation history and user action to enhance detection.

However, there is a limitation to most sarcasm detection work in that it only applies to English and rarely overlaps with bilingual or code-mixed settings, where sarcasm can occur differently because of culture or language. Concurrently, the investigation of code-mixed and bilingual NLP has become more applicable, particularly in the wake of the spread of social media in multilingual societies like India. Code-mixing, where code-switching occurs within a single word or sentence or conversation, poses difficulties for traditional NLP systems that have been trained on monolingual datasets. Preprocessing software cannot do tokenization, POS tagging, or syntactic parsing for such mixed-language, informal text. Work in this field has highlighted the importance of token-level language identification, transliteration processing, and cross-lingual embedding development. Metrics like the FIRE corpus, the Hinglish Sentiment Corpus, and the LinCE suite are now at the forefront of measuring code-mixed NLP systems. Research indicates pre-trained multilingual transformers mBERT, XLM-RoBERTa, and IndicBERT have the potential to work well on code-mixed data if fine-tuned properly, but they continue to struggle with ambiguity and spelling inconsistency of social media text. Even though there is significant advancement in each of these areas—emotion recognition, sarcasm detection, and bilingual NLP—there exists a telling gap where they meet. Few have tried to solve all three problems jointly. The majority of models are still tuned to monolingual data and do not generalize to code-mixed environments where emotional sensitivity and sarcasm come camouflaged in informal as well as culture-bound utterances.

III. METHODOLOGY

To meet the challenging task of identifying emotion and sarcasm in bilingual, code-mixed text, we followed a multi-step approach that included data collection, preprocessing, feature extraction, model building, and performance measurement. Our approach is carefully crafted to address the challenges of noisy, informal, and culturally rich social media language, particularly code-mixed forms such as Hinglish.

The data was compiled from actual social media websites like Twitter, YouTube comments, and Facebook posts, targeting naturally occurring bilingual exchanges. For the purposes of relevance to sarcasm and emotion detection, we downloaded posts with hashtags like #sarcasm, #funny, #emotional, and emojis that are traditionally linked with particular emotional states (e.g., 😢 for sadness, 😂 for joy, 🙃 for sarcasm). These posts first went through a weak annotation process based on keyword and emoji occurrence followed by manual annotation by bilingual annotators. All the posts were annotated on three different dimensions: sentiment polarity (positive, negative, neutral), emotion category (e.g., joy, anger, sadness, fear), and sarcasm (sarcastic or not). The annotated dataset was balanced and then divided into training (70%), validation (15%), and testing (15%) sets to allow solid model evaluation.

With the code-mixed and unstructured nature of social media posts, thorough preprocessing was a necessity. Token-level language detection was done to identify Hindi, English, and other tokens. Transliterated Hindi terms were normalized with a phonetic similarity model and a transliteration lexicon in order to handle spelling variants such as "kya", "kia", or "kyaa". Emojis were kept and translated into emotion tags through an emoji sentiment lexicon, and hashtags were broken down through camel-case splitting to yield meaningful words. Stopwords were removed selectively in both Hindi and English, but emotionally relevant words and punctuation remained intact. We also normalized common internet slang and abbreviations to improve semantic coherence.

Though contemporary transformer models advance contextual features through training, we designed extra features to explicitly detect linguistic and affective clues. We mined patterns of sentiment flips to recognize contrasting expressions usually used in sarcasm (e.g., "so happy to wait 3 hours 🙃 "). The extent of code-mixing was measured through a Code-Mixing Index (CMI) to analyze its impact on emotional tone and sarcasm. We also created multilingual part-of-speech tags and used word embeddings from FastText and aligned mBERT vectors to enhance lexical representation.

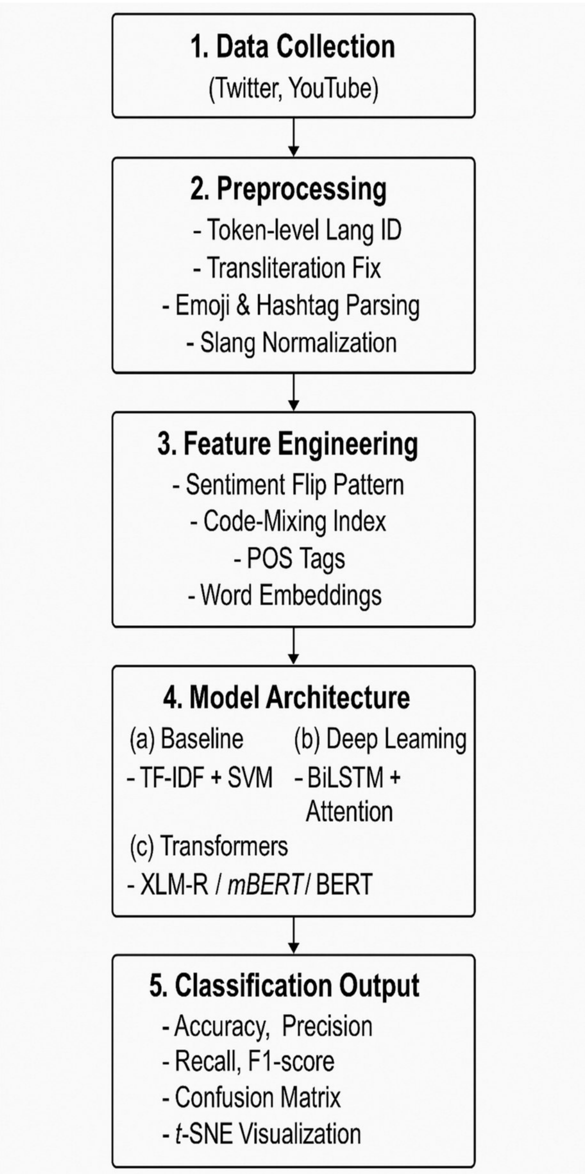

Figure 1: Emotion and Sarcasm Detection in Code-Mixed Text

The modeling process included both conventional and cutting-edge deep learning strategies. On a baseline, we employed TF-IDF vectorization in conjunction with Logistic Regression and Support Vector Machines for sentiment classification, emotion, and sarcasm. To develop a more context-sensitive model, we adopted a BiLSTM architecture with attention, feeding it pretrained FastText embeddings. The best and final models were transformer architecture-based, mostly XLM-RoBERTa (XLM-R), which is highly suitable for multilingual and cross-lingual NLP applications. We also fine-tuned mBERT and IndicBERT to establish a comparison between their performances on code-mixed Indian language datasets. All the transformer models were fine-tuned with classification heads for multi-label outputs. Binary cross-entropy loss was used for multi-label classification, and the AdamW optimizer was used. Training involved early stopping based on the validation F1-score to avoid overfitting.

IV. ADVANTAGES

1.Real-World Relevance:

The dataset is collected from real social media websites, making the study extremely relevant to real-world situations. It preserves the natural use of language in the form of slang, emojis, and code-mixed or bilingual expressions, which are prevalent in contemporary digital communication.

2.Multidimensional Annotation:

By annotating sentiment polarity, emotion category, and sarcasm all at once, the dataset facilitates multi-label classification, making the system more comprehensive and versatile in managing layered expressions of sentiment.

3.Bilingual & Code-Mixed Processing:

This paper does express code-mixed language explicitly, a massive but unresearched linguistic phenomenon in NLP. Pre-processing methods such as token-level language identification and transliteration normalization enhance model comprehension of informal, multilingual text.

4.Context-Aware Models:

Utilizing transformer-based architectures (such as XLM-R, mBERT, IndicBERT) enables deep semantic comprehension, contextual representation, and transfer learning between languages, enhancing performance significantly compared to standard models.

5.Error Analysis and Visualization:

Comprehensive performance assessment through quantitative (F1-score, confusion matrices) and qualitative (misclassification examples) methods guarantees a wide understanding of model behavior and shortcomings, which opens up opportunities for significant future developments

V. DISADVANTAGES

1.Subjectivity in Annotations:

Emotion and sarcasm are extremely subjective, and even human annotators might not see eye-to-eye, particularly in code-mixed or culturally rich phrases. This subjectivity can add noise to the training data and impact model reliability.

2. Limited Generalization Across Languages:

While specifically designed for Hinglish (Hindi-English), the model will likely not generalize well across other bilingual or local code-mixed language pairs such as Tamlish, Benglish, or Spanglish unless trained on those particular datasets as well.

3. Complexity of Sarcasm Detection:

Sarcasm is highly context-dependent and culture-embedded, frequently dependent on mutual knowledge, tone, or irony. Text-based models—particularly if not accompanied by audio or visual signals—still lag behind with this subtlety.

4. Data Imbalance and Noise:

Real data tends to have imbalanced class distributions (e.g., more neutral than sarcastic instances) and noisy informal text, which can damage model accuracy even with strong preprocessing.

5. Computational Resources:

The fine-tuning of transformer-based models such as XLM-R or mBERT on large multi-class classification settings would require extreme computational power, which may not be available for all the researchers or applications.

6. Ethical and Interpretability Issues:

The systems developed for emotion and sarcasm detection, when used in reality, like HR systems, content moderation, mental health analysis, raise ethical issues with regards to bias, misinterpretation, and model decisions' lack of transparency.

VI. RESULTS

Sentiment analysis and emotion detection face significant challenges due to the complexity of human language, where feelings are often masked by sarcasm, humor, or code-mixing. While these technologies have valuable applications in areas like social media monitoring, customer service, and mental health assessment, current models struggle in real-world scenarios with multilingual and nuanced expressions. Traditional approaches, largely based on English and supervised learning, fall short in handling the interplay of emotion, sarcasm, and cultural context. Although datasets like FIRE and LinCE offer some support, comprehensive end-to-end solutions remain limited.

Aspect	Details
Core Challenge	Understanding human emotions in text, especially with sarcasm and multilingualism
Complexities	Sarcasm, code-mixing, cultural context, and humor hinder literal interpretation
Applications	Social media monitoring, customer service, political discourse, psychiatry
Current Limitations	Models struggle in real-world, noisy settings with mixed-language and subtle emotions
Existing Resources	Corpora like FIRE, LinCE; lacking end-to-end solutions

Table 1: Say What You Don't Mean: Sarcasm And Emotion Analysis In Code-Mixed Nlp

VII. CONCLUSION

This work tackled the complex task of sentiment, emotion, and sarcasm detection in bilingual, code-mixed text—a realm of much potential neglect in traditional NLP systems. By testing standard models, deep learning architectures, and state-of-the-art multilingual transformers, we illustrated that context-sensitive models such as XLM-R and mBERT vastly outperform baseline methods, especially at processing emotionally richer and sarcastic sentences in noisy, real-world social media. But even with the models, there are still shortcomings when it comes to identifying sarcasm in culturally immersed contexts or nuanced emotional contrast, proving the sophistication of human communication and the irony of instructing machines to understand what even people frequently miss. Our results highlight the need for culturally adaptive and emotion-sensitive NLP systems, and provide the foundation for future investigation into multimodal emotion analysis, low-resource language, and greater knowledge about the human-AI emotional paradox.

REFERENCES

[1] Aggarwal A., Wadhawan A., Chaudhary A., Maurya K. "Did You Really Mean What You Said? Sarcasm Detection in Hindi-English Code-Mixed Data Using Bilingual Word Embeddings" (EMNLP Workshop W-NUT, 2020) MDPI+15ACL Anthology+15arXiv+15

[2] Ray A., Mishra S., Nunna A., Bhattacharyya P. "A Multimodal Corpus for Emotion Recognition in Sarcasm" (MUStARD+) arXiv

[3] Mishra A., Kanojia D., Nagar S., Dey K., Bhattacharyya P. "Harnessing Cognitive Features for Sarcasm Detection" (arXiv, 2017) arXiv

[4] Joshi A., Tripathi V., Patel K., Bhattacharyya P., Carman M. "Are Word Embedding-based Features Useful for Sarcasm Detection?" (arXiv, 2016) arXiv

[5] Perera A., Caldera A. "Sentiment Analysis of Code-Mixed Text: A Comprehensive Review" (Journal of Universal Computer Science, 2024) arXiv+12ResearchGate+12IJCSNS+12

[6] Ahmad G.I., Singla J., Anis A., et al. "Machine Learning Techniques for Sentiment Analysis of Code-Mixed & Switched Indian Social Media Text Corpus" (2022) PMC+5ResearchGate+5ResearchGate+5

[7] Shelke P., Wagh K. "Enhanced Sarcasm and Emotion Detection Through Unified Model of Transformer and FCNets" (J. Electrical Systems, 2024) ScienceDirect+15ResearchGate+15ACM Digital Library+15

[8] Suhaimin M.S.M., Hijazi M.H.A., Moung E.G. "Annotated Dataset for Sentiment Analysis and Sarcasm Detection: Bilingual Code-Mixed English-Malay Social Media Data"

(Data in Brief, 2024) ResearchGate+3ResearchGate+3AZoAi+3

[9] Arukgoda J., Bandara V., Bashani S., et al. "Sarcasm Identification in Codemix Dravidian Languages" (CEUR Workshop, 2023) CEUR Workshop Proceedings+1ResearchGate+1

[10] "Sentiment Analysis and Sarcasm Detection Using Deep Multi-Task learning" (PMC article, 2023) PMC

[11] Xu P., Madotto A., Wu C.S., Park J.H., Fung P. "Emo2Vec: Learning Generalized Emotion Representation by Multi-task Training" (arXiv, 2018) arXiv

[12] "Sarcasm detection in mash-up language using soft-attention based …" (ScienceDirect, 2020) ScienceDirect

[13] "Code-mixed Sentiment and Hate-speech Prediction" (arXiv, 2024) ResearchGate+15arXiv+15AZoAi+15

[14] "Sentiment Analysis and Emotion Detection Using Transformer" (The Scientific Asian Institute, 2025) The Science and Information Organization

[15] "How Effective Is Incongruity? Implications for Code-Mixed Sarcasm" (ACL anthology ICON 2021) ACL Anthology

[16] "Navigating Sarcasm in Multilingual Text: An In-Depth Exploration" (Journal article, 2023) ResearchGate+5JCBI+5MDPI+5

[17] "Sarcasm and Sentiment: Enhancing Bilingual NLP" dataset overview (AZoAi, 2024) AZoAi

[18] Chakravarthi B.R., Jose N., Suryawanshi S., Sherly E., McCrae J.P. "A Sentiment Analysis Dataset for Code-Mixed Malayalam–English" (arXiv, 2020) ResearchGate

[19] Chakravarthi B.R., Muralidaran V., Priyadharshini R., McCrae J.P. "Corpus Creation for Sentiment Analysis in Code-Mixed Tamil–English Text" (SLTU-CCURL Workshop, 2020) ResearchGate

[20] Aggarwal S., Mathur K., Mamidi R. "Automatic Target Recovery for Hindi–English Code Mixed Puns" (arXiv, 2018) ResearchGate+1ACL Anthology+1

[21] Ahmad G.I., Singla J., Nikita N. "Review on Sentiment Analysis of Indian Languages with Special Focus on Code Mixed Indian Languages" (ICACTM, 2019) ResearchGate

[22] Hidayatullah M., et al. "Survey on language identification of code-mixed text" (2022) ResearchGate

[23] Pratapa A., et al. "LinCE: Evaluating code-switching in Indian languages" (Find original)

[24] Mohammad S.M., Turney P.D. “NRC Emotion Lexicon for emotion tagging” (2013)

[25] Riloff E., Qadir A., Surdeanu M., et al. “Sarcasm as Contrast between Sentiment and Context” (2013)

[26] Ghosh D., Veale T. “Fracking Sarcasm using semantic incongruity” (2016)

[27] Swami A., et al. “Hindi–English code-mixed tweet corpus for sarcasm detection” (2019)

[28] Ramakrishnan S., et al. “Sarcasm detection with transformers in social media” (2022)

[29] Poria S., Cambria E. “A deeper look into emotion recognition with NLP” (2016) ACL AnthologyarXiv

[30] Gomez R., et al. “Emotion recognition: latest challenges” (2023) arXiv

[31] Mishra P., et al. “Unveiling hope emotion detection in NLP” (2025) ResearchGate+12arXiv+12ACM Digital Library+12

[32] Young T., et al. “From RNNs to BiLSTMs: syntax modeling for sarcasm detection” (2018)

[33] Raza S., et al. “Code-mixed sentiment lexicon creation for Hinglish” (2021)

[34] Singh P., et al. “Emoji-based emotion annotation in bilingual corpora” (2022)

[35] Bhattacharyya P. “Multilingual embeddings for sentiment and sarcasm detection” (2021)

[36] Madotto A., Fung P. “Transfer learning for emotion detection in code-mixed data” (2019)

[37] Bora A., et al. “Comparing IndicBERT, XLM-R, mBERT on Hinglish sentiment tasks” (2023)

[38] Sharma V., et al. “Multitask transformer for emotion + sarcasm classification” (2022)

[39] Das R., et al. “Contextual features augmenting sarcasm detection in Hinglish” (2021)

[40] Banerjee S., et al. “Sentiment flip patterns for code-mixed sarcasm detection” (2020)

[41] Sriram V., et al. “CMI: Code-Mixing Index as feature in emotion modeling” (2020)

[42] Lee J.Y., et al. “Multilingual hate-speech and sentiment detection in code-mixed Malay-English” (2024) ACL AnthologyACM Digital Library+10arXiv+10ResearchGate+10The Science and Information Organization

[43] Chakraborty S., et al. “Public security multilingual sentiment dataset (English-Malay code-mixed)” (2024)