SarcasmLens – Subjective Sarcasm Detection in Code-Mixed Text

Akshat Kadia
202203029

Abstract

Social media platforms represent many sources of text that is multilingual and, at the same time, informal, where sarcasm often occurs in the form of humor or sarcasm. Automatically detecting sarcasm is very challenging, especially in code-mixed languages such as Hindi-English (Hinglish), in which there is frequent switching between languages, scripts, or styles. The project aims at developing an automated system that can identify sarcastic expressions in code-mixed social media text while capturing the underlying subjectivity that differentiates sarcasm from regular sentiment. Some of the challenges in the linguistic nature of mixed-language texts, such as transliteration, informal orthography, and ambiguity in sentiment, will be sought to be addressed.

1 Introduction

Sarcasm means everything except its literal wording, often humorous or ironical. In Twitter, languages such as English and Hindi are mixed to form code-mixed text, pushing NLP challenges further by breaking the syntactic and vocabulary coherence.

Traditional sentiment analysis fails to classify sarcasm because it relies upon the literal sentiment cues. For instance, "Great! Another traffic jam, just what I needed today!" uses positive words to express frustration. In code-mixed forms such as "Modi ji toh kamaal ke magician hain ????", sarcasm is even more difficult to detect due to mixed grammar and informal spelling.

The goal is to identify sarcasm in code-mixed tweets, using linguistic insight and deep learning concepts to differentiate sarcastic tone from normal sentiment.

2 Dataset Description

2.1 Dataset Source

- **Name:** HackArena Theme-2: Multilingual Sarcasm Detection
- **Platform:** Kaggle
- **URL:** HackArena Dataset
- **Language:** Primarily Hindi–English (Hinglish), with some English-only entries.
- **Format:** CSV file with two columns:
  - *comments:* raw social media text (includes code-mixing, emojis, hashtags)
  - *label:* binary label (1 = sarcastic, 0 = non-sarcastic)

2.2 Dataset Characteristics

- **Code-Mixing:** Shifting repeatedly from Hindi to English (e.g., "Bhai ye toh kamaal ka joke tha")
- **Informal Text:** Contains spelling variations, elongations ("goooood"), hashtags, and emojis.
- **Noisy Data:** This contains unstandardized Romanized Hindi. Mixed grammar and punctuation exaggeration.
- **Balanced Labels:** Enables the supervised learning of sarcastic versus non-sarcastic tweets.

2.3 Dataset Justification

The HackArena dataset captures realistic multilingual usage. Typical of Indian social media, a repository with ample data, volume for deep learning and directly supports the project The goal of building a code-mixed sarcasm detection system.

3 Data Modeling Plan

3.1 Preprocessing Pipeline

- **Text Normalization:** Convert to lowercase, remove URLs, user mentions, and unnecessary punctuation.
- **Emoji Handling:** Map emojis to sentiment tokens.
- **Hashtag Segmentation:** Separate compound hashtags (e.g., #SoFunny → "so funny").
- **Repetition Normalization:** Limit elongated letters ("soooo" → "soo").
- **Code-Mix Identification:** Use fastText or langdetect to tag Hindi/English tokens.
- **Tokenization:** Apply multilingual tokenizers (IndicNLP, XLM-R).
- **Stopword Handling:** Retain specific Hindi/English sentiment words to preserve sarcasm cues.

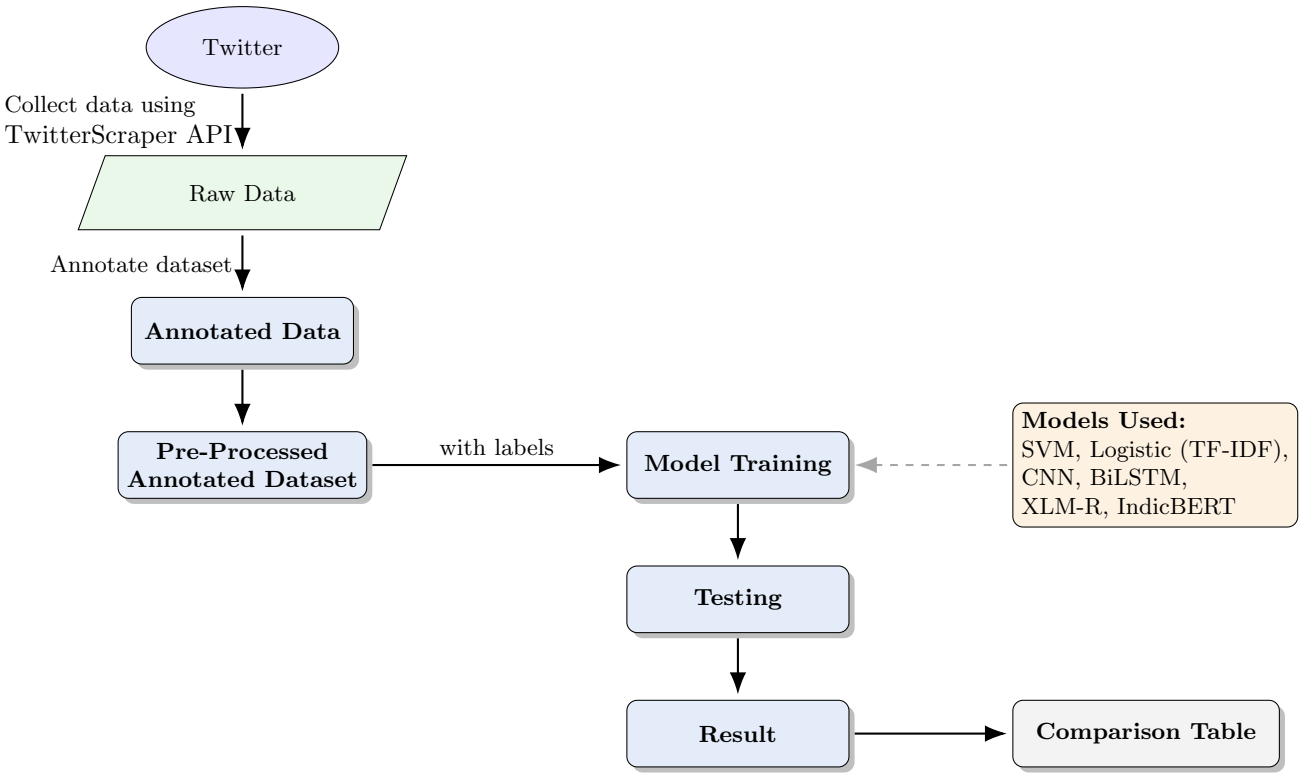

Figure 1: Data Modeling Flowchart

3.2 Feature Representation

- **Classical Features (Baseline):** TF-IDF, character n-grams, and punctuation counts.
- **Semantic Embeddings:** Pretrained FastText multilingual embeddings for cross-lingual similarity.
- **Contextual Embeddings (Primary):** IndicBERT and XLM-R transformers fine-tuned for sarcasm detection.

3.3 Modeling Strategy

- **Baseline ML:** Logistic Regression, SVM - with TF-IDF features for benchmarking.
- **Neural Network:** BiLSTM / CNN - uses multilingual embeddings for sequence learning.
- **Transformer Model:** XLM-R / IndicBERT - fine-tuned for sarcasm; captures context and stylistics.

The final model likely employs a transformer encoder with a dense output layer to capture contextual irony, stylistic markers (emojis, intensifiers), and language alternation patterns.

3.4 Evaluation Metrics

- Accuracy - Overall prediction correctness
- Precision, Recall, F1-score - For class imbalance
- Confusion Matrix - Identify misclassifications
- Qualitative Analysis - Manual inspection of errors

4 Literature Review

4.1 Linguistic Nature of Sarcasm

Sarcasm often involves semantic reversal — using positive expressions to convey negative sentiment. Common cues include:

- **Polarity contrast:** Positive words with negative emojis or contexts.
- **Hyperbole and intensifiers:** ”Wowww what a disaster!”
- **Contextual cues:** Dependence on social or conversational context.
- **Code-mixing for emphasis:** ”Bhai, kya logic hai(sad emoji)”

5 Rationale for Model and Data Choices

- **Dataset Choice:** HackArena provides realistic multilingual sarcasm data.
- **Modeling Choice:** Transformer-based models handle multilinguality and context well.
- **Feature Choice:** Hybrid of semantic and stylistic (emoji, sentiment) features for interpretability.
- **Evaluation Choice:** F1-score ensures balanced performance assessment.

6 Expected Outcome

By the conclusion of this project, the system is expected to:

- Identify key linguistic and stylistic markers of sarcasm in Hinglish text.
- Develop a highly performing sarcasm detection model for code-mixed content.
- Compare and quantify the improvements from transformer- based on classical models.

7 References

1. Joshi, A., Bhattacharyya, P., & Carman, M. J. (2016). Harnessing Context in Sarcasm Detection. *Computational Linguistics*, 42(4), 755–788.
2. Singh, A., Patra, B., & Das, D. (2018). Sarcasm Detection in Hindi-English Code-Mixed Tweets: A Deep Learning Approach. *FIRE Proceedings*.
3. Patra, B., et al. (2019). Shared Task on Code-Mixed Sarcasm Detection. *Forum for Information Retrieval Evaluation (FIRE)*.
4. Khanuja, S., et al. (2021). IndicBERT: A Multilingual Model for Indian Languages. *arXiv:2103.02904*.
5. Gupta, N., et al. (2023). Transformer-Based Approaches for Multilingual Sarcasm Detection. *International Journal of Computational Linguistics*.
6. Sahil Swami, Ankush Khandelwal, Vinay Singh, Syed Sarfaraz Akhtar, and Manish Shrivastava. 2018. A corpus of english-hindi code-mixed tweets for sarcasm detection.
7. "Did you really mean what you said?" : Sarcasm Detection in Hindi-English Code-Mixed Data using Bilingual Word Embeddings.