ISBN: 978-981-94-4198-3
2025 the 15th International Workshop on Computer Science and Engineering (WCSE 2025)
doi: 10.18178/wcse.2025.06.013

Filipino Genz Slang Sentiment Analyzer using BERT

Ria A. Sagum[+], Mark Ciedrick A. Ramos, Diana Rose A. Certeza, Alijah Czareen A. Andres, Aaronn Daphne M. Gatchalian
Polytechnic University of the Philippines, Philippines

Abstract: The rapid evolution of online communication has led to the emergence of Filipino Gen Z slang, a dynamic and evolving linguistic phenomenon. Traditional sentiment analysis models often struggle with understanding these slang expressions, necessitating the development of a more specialized approach. This study explores the application of BERT-based model to recognize and analyze sentiment in Filipino Gen Z slang. The model is fine-tuned using a curated dataset of slang expressions to improve accuracy in sentiment classification. Additionally, the research contributes to the advancement of natural language processing (NLP) in low-resource languages and offers insights into modern Filipino digital communication. The results demonstrate the effectiveness of transformer-based architecture in capturing the nuances of Filipino Gen Z slang, providing a more context-aware sentiment analysis tool.

Keywords: GenZ slang, sentiment analysis, BERT, NLP, deep learning, transformer models.

1. Introduction

Cyberworld has been evolving throughout the era, shaping and reflecting the societal culture and norms, which also brings out generational differences. Especially in social media, through these online communities that promote group interaction, people of all ages and demographics are enabled to exchange thoughts, passions, and viewpoints. The Generation Z— those born 1995 and 2010 and are also called Genzers or the GenZ [1]—, a group recognized for its digital nativeness and inventiveness, often spend time on social media platforms for connectedness and enjoyment. Because it was widely accessible worldwide, this increased their ability to influence and convey their thoughts and opinions [2]. Given the ever-changing nature of social media, GenZ also paves way to a whole new level of perspective when dealing with the platform, numerous new words are formed and are being introduced every day.

The evolution of language has been influenced by human ingenuity, which gave rise to slang as a prevalent and accepted form of communication in contemporary society [3]. Internet slang tends to create confusion for people unfamiliar with it, and the rapid growth and creation of these expressions used by the members of the Filipino Gen Z tends to add to this. This includes words such as “accla”, “tita/tito”, “dasurv”, “korique”, and “charot/chariz” which also contributes to the generation gap in understanding these slangs used in the cyber space. The rapid evolution of GenZ slang and the underlying cultural changes associated with such terms add difficulty and confusion to understanding it [4]. In connection with this, sentiment analysis issues manifested for those sentiments that ended in punctuation, and that some words were incorrectly tagged by their program resulting in misclassification. GenZ Slangs are known to include different linguistic patterns, which includes multiple punctuations [5].

Sentiment analysis is increasingly used in social media monitoring and business to gauge public opinion and customer feedback. However, current tools often struggle with the nuances of Gen Z slang, causing inaccurate interpretations. Understanding Filipino GenZ slang sentiment on social media is particularly challenging due to its intricacy and rapid change, and the limited research available [6]. This study contributes to developing better AI to interpret these sentiments, improving sentiment analysis for practical use.

[+] Corresponding author.
E-mail address: rasagum@pup.edu.ph;

2. Literature Review

2.1. Sentiment Analysis and Slang

Sentiment analysis, also known as subjectivity analysis, opinion mining, or emotion AI, is a technique in Natural Language Processing that extracts meaningful patterns from large text datasets to assess thoughts, opinions, and emotions rather than reasoning. It evaluates sentiments expressed in various sources like social media, reviews, blogs, and speeches, focusing on entities such as products, services, topics, and organizations. Typically, sentiment analysis involves three key elements: opinions or emotions, which can be qualitative (e.g., joy, anger, surprise) or quantitative (e.g., ratings); the subject of discussion, where multiple aspects of an entity can be evaluated, and the opinion holder, referring to the individual expressing the sentiment [7].

In recent years, sentiment analysis has become increasingly popular in social media and business [8]. The rule-based or lexicon-based approach is commonly used, relying on a predefined dictionary of words with manually assigned valence scores. These algorithms match words from the lexicon to those in the text, summing or averaging the scores to determine the overall sentiment of a sentence or document. While this method is fast, it may struggle with certain tasks, as the polarity of words can vary depending on the context, which may not be captured by the static dictionary [7].

Sentiment analysis typically involves several key stages: data collection, pre-processing (including tokenization, stop word removal, and stemming), feature extraction (converting text to feature vectors for nuanced sentiment detection), sentiment classification using algorithms, and finally, polarity determination (positive, negative, or neutral) [9].

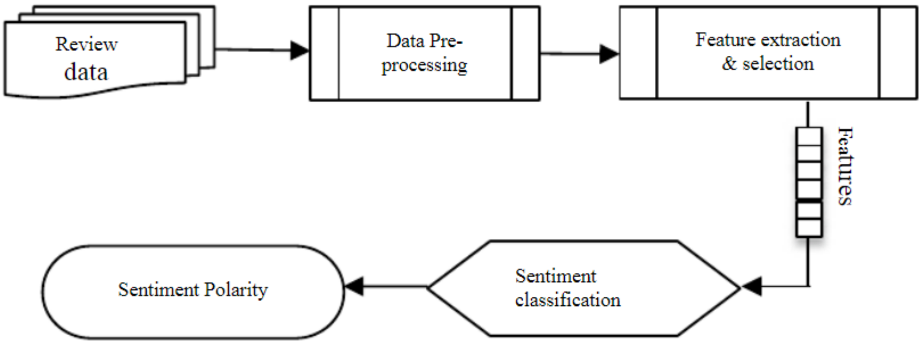

Fig. 1: Sentiment Analysis Process

Human creativity has driven language evolution, resulting in the widespread acceptance of slang in modern communication [10], [11]. Slang, a form of informal language specific to social groups, presents a unique linguistic challenge, being widely used yet difficult to precisely define [12]. Slang embodies the adaptability of language in the digital age, acting as a crucial tool for casual online interaction. Furthermore, its use promotes a feeling of belonging among peers, strengthening social connections within digital environments [13].

There is a critical need to understand slang [14], as its sentiment can be difficult for those accustomed to formal language, and its presence can negatively impact sentiment analysis accuracy. Similarly, the challenge of processing noisy slang terms in social media sentiment analysis notes that both formal and informal language can introduce ambiguity [6], [14]. This is due to slang's context-dependent meanings, which require accurate interpretation for nuanced sentiment detection. Furthermore, slang’s deviation from standard words can lead to language processing errors [15].

2.2. BERT in Sentiment Analysis

The development of deep learning models, especially transformer-based models such as BERT, has led to significant progress in sentiment analysis. These models are well-suited to handle the informal and dynamic language of social media as they have demonstrated superior performance in capturing context and semantic relationships in text [16], [17]. BERT, despite its promising results in NLP tasks, encounters several challenges, particularly when dealing with non-standard language. Specifically, its reliance on formal language training leads to difficulties in processing slang, often resulting in misclassification of emotion-bearing text as neutral due to the lack of strong emotional cues or polarized adjectives that BERT models typically rely on for sentiment classification [18]Additionally, BERT's monolingual design necessitates significant fine-tuning to effectively handle multilingual language data, such as code-switching languages like Taglish, where its inherent limitations in capturing cross-lingual nuances become apparent [19].Finally, BERT's performance is negatively impacted by imbalanced datasets, which can skew classification results [20].

3. Methodology

3.1. Research Design

This study employed an experimental design to evaluate the sentiments of Filipino Gen Z slang. The methodology involved fine-tuning a BERT model to analyze the linguistic nuances of Filipino Gen Z slang. Reddit content was maintained in its original form, with a focus on extracting and analyzing specific linguistic features to understand their effect on sentiment and emotion expression within online discourse. This approach provided empirical insights into the influence of linguistic evolution on digital communication.

3.2. System Architecture

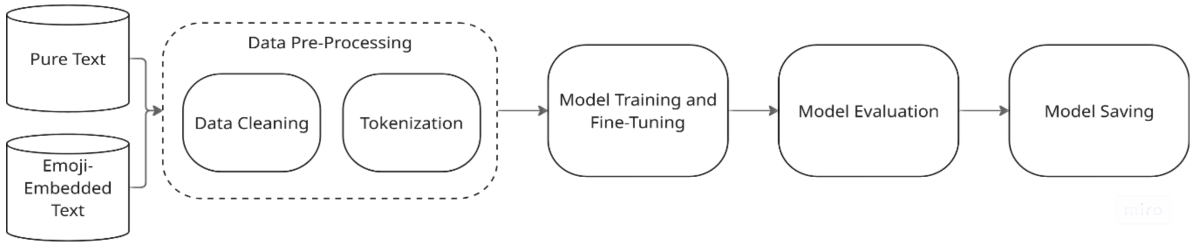

Fig. 2: System Architecture

There are two primary data sources from Reddit: pure text and emoji-embedded text. This raw data first undergoes Data Pre-Processing, which includes Data Cleaning (removing noise like hashtags, mentions, links, punctuation, special characters except emojis, and normalizing repeated characters) and Tokenization using the BERT-uncased tokenizer. The pre-processed data is then split into training (80%), validation (10%), and testing (10%) sets for Model Training and Fine-Tuning of a pre-trained BERT model to understand sentiment in Filipino Gen Z slang. The trained model's performance is assessed in the Model Evaluation stage using standard sentiment analysis metrics. Finally, the validated model is saved during Model Saving for future deployment.

3.3. Dataset

This study primarily used publicly available posts and comments from Filipino subreddits such as but not limited to r/Philippines, r/PHGamers, and r/sanaall on Reddit, accessed via the Reddit public API, due to its accessibility—compared to other platforms that restricts users to gather data efficiently—,and due to the diverse Filipino communities. Employing a non-probabilistic purposive sampling technique, data collection focused on identifying Filipino Gen Z slang, guided by the morphological structures identified in prior research (Grandez et al., 2023; Jeresano & Carretero, 2022; Gime & Macascas, 2020; Nacorda & Garma, 2022). The collected data, comprising 2400 items (1200 text-only, 1200 emoji-embedded) split into training (80%), validation (10%), and testing (10%) sets, underwent preprocessing, including cleaning (removal of hashtags, mentions, links, punctuation, special characters (except emojis), and repeated characters) and tokenization using the BERT-uncased tokenizer. A limitation of this study is that text validation was primarily conducted by the Gen Z researchers, potentially introducing subjectivity, and data collection focused on posts containing Gen Z slang, not author's generational identity.

3.4. Evaluation Metrics

The model's performance in sentiment classification was evaluated using a confusion matrix. Precision, recall, and F1-score were calculated from this matrix, where:

- TP (True Positive): Positive sentiment correctly identified.
- TN (True Negative): Negative sentiment correctly identified.
- FP (False Positive): Negative sentiment incorrectly classified as positive.
- FN (False Negative): Positive sentiment incorrectly classified as negative

3.5. Ethical Considerations

Ethical considerations were addressed by ensuring that data collection from Reddit focused on public posts and comments, thus not requiring individual user consent. Anonymity and confidentiality measures were implemented to safeguard user privacy, and steps were taken to avoid harm and bias in the data collection and analysis.

4. Results and Discussion

Table 1: Confusion matrix for sentiment classification on plain texts

SENTIMENT			
ACTUAL	PREDICTED		TOTAL
	Positive	Negative	
Positive	561	39	600
Negative	50	550	600

Table 1 shows the confusion matrix for plain text sentiment classification, distinguishing between positive and negative sentiments. Of the 600 positive sentiments, 561 were correctly classified, and 39 were misclassified as negative. Similarly, 550 of the 600 negative sentiments were correctly classified, with 50 misclassified as positive. These results indicate the model's strong performance in plain text sentiment classification, with minor misclassifications suggesting potential improvements. The model's plain text performance was further evaluated using precision, recall, and F1-score (Table 3).

Table 2: Confusion matrix for sentiment classification on plain texts

SENTIMENT			
ACTUAL	PREDICTED		TOTAL
	Positive	Negative	
Positive	551	49	600
Negative	127	473	600

Table 2 presents the confusion matrix for emoji-embedded text. Of the 600 positive sentiments, 551 were correctly classified, and 49 were misclassified as negative. For the 600 negative sentiments, 473 were correctly classified, with 127 misclassified as positive. While showing high true positive (551) and true negative (473) counts, the model had more difficulty distinguishing negative sentiments in emoji-embedded text, indicated by the increased false positives (127) compared to plain text classification.

Table 3: Performance of the model in analyzing sentiments

	Precision	Recall	F1-Score
Plain Texts	0.92	0.94	0.93
Emoji-Embdedded	0.81	0.92	0.86
Combined	0.86	0.93	0.89

The model's precision, recall, and F1-score are summarized in Table 3. Plain text classification demonstrated high performance (precision: 0.92, recall: 0.94, F1-score: 0.93). Emoji-embedded text showed reduced precision (0.81), indicating more false positives, though recall remained high (0.92), with an F1-score of 0.86.

The overall performance was precision 0.86, recall 0.93, and F1-score 0.89. These results suggest strong sentiment detection, particularly in recall, but also indicate that emojis complicate accurate sentiment classification, requiring further refinement for nuanced slang.

5. Conclusion

In conclusion, this study has demonstrated the potential of a BERT-based model for effectively recognizing and analyzing sentiment within the complex and evolving landscape of Filipino Gen Z slang. The fine-tuning process enabled the model to capture the nuances inherent in this informal language, showcasing the strength of transformer-based architectures in providing context-aware sentiment analysis. The results indicate that the model can reliably classify sentiments in plain text, although challenges remain in accurately processing emoji-embedded text, suggesting areas for further refinement. Overall, this research contributes to the advancement of natural language processing techniques for low-resource languages and provides valuable insights into the sentiment dynamics of modern Filipino digital communication.

6. Recommendation

Several improvements are needed to further advance sentiment analysis of Filipino Gen Z slang. Emoji handling requires attention; emoji-specific embeddings could improve the model's lower precision with emoji-rich text. Dataset expansion is also important to enhance robustness, incorporating more diverse slang, linguistic patterns, and sources from various platforms and Gen Z subgroups. Moreover, future work should develop better contextualization techniques to address slang ambiguity, potentially using pragmatic or discourse-level analysis. Given the prevalence of code-switching, particularly Taglish, multilingual model development is crucial. Finally, exploring real-world applications, such as customer feedback or mental health support systems, would validate the research's practical contributions.

7. References

[1] E. M. Jeresano and M. D. Carretero, “Digital Culture and Social Media Slang of Gen Z,” 2022. [Online]. Available: https://www.deped.gov.ph

[2] D. Singh and N. Guruprasad, “Impact of Social Media on Youth,” 2019. doi: http://dx.doi.org/10.2139/ssrn.3506607.

[3] R. Maulidiya, S. E. Wijaya, C. Mauren, T. Pra Adha, M. Glorino, and R. Pandin, “Language Development of slang in the Younger Generation in the Digital Era.” Accessed: Apr. 02, 2025. [Online]. Available: https://doi.org/10.31219/osf.io/xs7kd

[4] S. Poria, D. Hazarika, N. Majumder, and R. Mihalcea, “Beneath the Tip of the Iceberg: Current Challenges and New Directions in Sentiment Analysis Research,” 2020. [Online]. Available: https://paperswithcode.com/task/sentiment-analysis.

[5] R. A. Sagum*, M. L. Navarro, and A. Jasper E, “EMOSIS Sentiment Analysis on Tweets with Emotion and Intensity Level Recognition Considering Ending Punctuation Marks,” International Journal of Recent Technology and Engineering (IJRTE), vol. 8, no. 4, pp. 10289–10293, Nov. 2019, doi: 10.35940/ijrte.D4518.118419.

[6] T. Kolajo, “Sentiment Analysis on Naija-Tweets,” 2019.

[7] M. Lamba and M. Madhusudhan, “Sentiment Analysis,” in Text Mining for Information Professionals, Springer International Publishing, 2022, pp. 191–211. doi: 10.1007/978-3-030-85085-2_7.

[8] N. A. Semary, W. Ahmed, K. Amin, P. Pławiak, and M. Hammad, “Improving sentiment classification using a RoBERTa-based hybrid model,” Front Hum Neurosci, vol. 17, 2023, doi: 10.3389/fnhum.2023.1292010.

[9] D. Sharma, M. Sabharwal, V. Goyal, and M. Vij, “Advances in Intelligent Systems and Computing 1045 First International Conference on Sustainable Technologies for Computational Intelligence Proceedings of ICTSCI 2019.” [Online]. Available: http://www.springer.com/series/11156

[10] K. Sadia and S. Basak, “Sentiment Analysis of COVID-19 Tweets: How Does BERT Perform?,” 2021, pp. 407–416. doi: 10.1007/978-981-16-0586-4_33.

[11] R. Maulidiya, S. E. Wijaya, C. Mauren, T. Pra Adha, M. Glorino, and R. Pandin, “Language Development of slang in the Younger Generation in the Digital Era.”

[12] F. W. Nuraeni, J. Pahamzah, U. Sultan, and A. Tirtayasa, “AN ANALYSIS OF SLANG LANGUAGE USED IN TEENAGER INTERACTION.”

[13] J. Manurung, M. Helentina Napitupulu, and H. Simangunsong, “Exploring the Impact of Slang Usage Among Students on WhatsApp: A Dig-ital Linguistic Analysis,” vol. 11, no. 2, pp. 153–169, 2022, [Online]. Available: https://journals.ristek.or.id/index.php/jiph/index

[14] “Improving Sentiment Analysis of Shopee Reviews with Informal Language and Slang,” Journal of Logistics, Informatics and Service Science, vol. 11, no. 3, Apr. 2024, doi: 10.33168/jliss.2024.0311.

[15] P. Nandwani and R. Verma, “A review on sentiment analysis and emotion detection from text,” Dec. 01, 2021, Springer. doi: 10.1007/s13278-021-00776-6.

[16] B. Li, X. Liu, and R. Zhang, “Employing the BERT model for sentiment analysis of online commentary,” Applied and Computational Engineering, vol. 32, no. 1, pp. 241–247, Jan. 2024, doi: 10.54254/2755-2721/32/20230218.

[17] B. V Pranay Kumar and M. Sadanandam, “A Fusion Architecture of BERT and RoBERTa for Enhanced Performance of Sentiment Analysis of Social Media Platforms. A Fusion Architecture of BERT and RoBERTa for Enhanced Performance of Sentiment Analysis of Social Media Platforms.” [Online]. Available: https://ssrn.com/abstract=4455231

[18] A. Chiorrini, C. Diamantini, A. Mircoli, and D. Potena, “Emotion and sentiment analysis of tweets using BERT,” 2021. [Online]. Available: https://www.researchgate.net/publication/350591267

[19] F. A. Acheampong, H. Nunoo-Mensah, and W. Chen, “Transformer Models for Text-based Emotion Detection: A Review of BERT-based Approaches.” [Online]. Available: https://www.researchgate.net/publication/348740926

[20] L. Luo and Y. Wang, “EmotionX-HSU: Adopting Pre-trained BERT for Emotion Classification,” Jul. 2019, [Online]. Available: http://arxiv.org/abs/1907.09669