IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, January 2026
Impact Factor: 7.67

Comparative Analysis of Word Embeddings on Transformer Model for Emotion Recognition in Indic Code-Mixed Hinglish

Pragati Mulik[1] and Dr. S. S. Sonawane[2]
M.Tech. Student, Department of Computer Engineering[1]
Professor, Department of Computer Engineering[2]
SCTR's Pune Institute of Computer Technology, Pune,India

Abstract: *Hinglish, a code-mixed blend of Hindi and English, has become increasingly common in digital communication across platforms such as WhatsApp, Instagram, and Twitter. Its informal grammatical structure, transliterated Hindi tokens in Roman script, and frequent language switching pose significant challenges for traditional NLP systems trained on monolingual corpora. This paper presents a comparative approach for Hinglish emotion recognition using four embedding-classifier combinations, namely Skip-gram + LLaMA, CBOW + LLaMA, BERT + LLaMA, and SBERT + LLaMA. A dataset of 16,000 Hinglish sentences annotated with seven emotion categories-joy, anger, sadness, fear, love, surprise, and neutral was used for experimentation. A specialized preprocessing pipeline was developed to address transliteration inconsistencies and spelling variations. The models were evaluated using accuracy and weighted F1-score. Among the methods tested, the CBOW + LLaMA model achieved the highest performance, followed closely by the BERT + LLaMA model. The study highlights the suitability of context-preserving embeddings for code-mixed Indic text and supports the development of practical emotion-aware NLP systems for multilingual Indian users.*

Keywords*:* Hinglish; Code-Mixed Text; Emotion Recognition; Multilingual NLP; Deep Learning; Transformer Models; LLaMA

DOI: 10.48175/IJARSCT-31012