

# Chapter 12

## Cloud-Based Offensive Code Mixed Text Classification Using Hierarchical Attention Network

**Durga Karthik**

 <https://orcid.org/0000-0003-3199-8814>  
SASTRA University, India

**Rajeswari Natarajan**

SASTRA University, India

**R. Bhavani**

SASTRA University, India

**D. Rajalakshmi**

SASTRA University

### ABSTRACT

*The use of mixed language in social media has increased and the need of the hour is to detect abusive and offensive content. Hierarchical attention network (HAN) is employed for classifying offensive content both at word and sentence level. Data from Thinkspeak cloud tweets containing annotated Tamil and English text is used as a training set for the HAN model. The attention mechanism captures the significance from both word and sentence levels. Cross-entropy loss function and backpropagation algorithm in the model classify offensive code-mixed text with an accuracy of 0.58. The above model can be employed for classifying other mixed language text too.*

DOI: 10.4018/979-8-3693-1694-8.ch012

## **INTRODUCTION**

Classifying offensive code in social media content is difficult because of the mixed language use. Social media text are difficult for analysis as they do not follow grammar, are non-vocabulary, presence of slang usage and with native contents in English. Sentiment analysis on codemixed data involves categorizing it as positive or negative. LSTM based model for the above task has yielded a good f1 score and rank (Chaturanga & Ranathunga, 2021). Preprocessing of code mixed text includes splitting, stopword removal and stemming. Machine learning techniques can be employed for classification and XLM classifier outperforms BERT for YouTube content containing English and Malayalam text (Kazhuparambil & Kaushik, 2020).

Ensemble methods on various ML algorithms can be employed for code mixed sentiment analysis. An ensemble voting classifier on Dravidian languages with English code for message level polarity yielded nearly 0.6 accuracy (Hidayatullah et al., 2023). Embedded models Capsule with BIGRU outperforms BERT for English language and XLM\_R for Sinhala code mixed with English. The above recurrent model for baselining followed by improvements such as regularization, integration and extraction on the text. As a final step a capsule network was formed to classify Sinhalese text and obtained an accuracy of 0.82 (Kalaivani & Thenmozhi, 2020). Tamil language code mixed with English is very common and sentiment analysis using Deep learning for predicting tags and for character embedding had a better f1 score than other machine learning algorithms on the same data set (Kazhuparambil & Kaushik, 2020).

Indonesian code mixed with English and Japanese tweets were tuned with BERT for word level identification of the language and is a reliable model even for sub word level identification with a good precision and recall (Kumar Mahata et al., 2021). Sentiment analysis and identification of offensive content on Dravidian languages using Adaptive BERT has yielded 0.79 accuracy (Mariyam et al., 2023). Malayalam with English code mixed data classification for both sentiment and offensive content using Deep learning models has yielded an accuracy of 0.76 (Rashmi et al., n.d.). HAN has an accuracy of 0.96 in classifying long document in both sentence and word level (Shanmugavadivel et al., 2022). HAN helps in optimal feature selection for text classification using Invasive weed tunicate method (Singh et al., 2023) and has lead to Improved Invasive weed optimization technique for text classification (Singh et al., 2022).

12 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/chapter/cloud-based-offensive-code-mixed-text-classification-using-hierarchical-attention-network/341004](http://www.igi-global.com/chapter/cloud-based-offensive-code-mixed-text-classification-using-hierarchical-attention-network/341004)

## Related Content

---

### A Heuristic Meta Scheduler for Optimal Resource Utilization and Improved QoS in Cloud Computing Environment

R. Jeyarani and N. Nagaveni (2012). *International Journal of Cloud Applications and Computing* (pp. 41-52).

[www.irma-international.org/article/heuristic-meta-scheduler-optimal-resource/64634](http://www.irma-international.org/article/heuristic-meta-scheduler-optimal-resource/64634)

### A Novel Multi-Secret Sharing Approach for Secure Data Warehousing and On-Line Analysis Processing in the Cloud

Varunya Attasena, Nouria Harbi and Jérôme Darmont (2019). *Cloud Security: Concepts, Methodologies, Tools, and Applications* (pp. 483-506).

[www.irma-international.org/chapter/a-novel-multi-secret-sharing-approach-for-secure-data-warehousing-and-on-line-analysis-processing-in-the-cloud/224590](http://www.irma-international.org/chapter/a-novel-multi-secret-sharing-approach-for-secure-data-warehousing-and-on-line-analysis-processing-in-the-cloud/224590)

### Analyzing Virtualization Vulnerabilities and Design a Secure Cloud Environment to Prevent from XSS Attack

Nitin Nagar and Ugrasen Suman (2016). *International Journal of Cloud Applications and Computing* (pp. 1-14).

[www.irma-international.org/article/analyzing-virtualization-vulnerabilities-and-design-a-secure-cloud-environment-to-prevent-from-xss-attack/144185](http://www.irma-international.org/article/analyzing-virtualization-vulnerabilities-and-design-a-secure-cloud-environment-to-prevent-from-xss-attack/144185)

### Towards Connected Governance: Citizens' Use of Web 2.0 in Nigeria

Kemi Ogunsola and Temilade Precious Olojo (2021). *Web 2.0 and Cloud Technologies for Implementing Connected Government* (pp. 68-94).

[www.irma-international.org/chapter/towards-connected-governance/259735](http://www.irma-international.org/chapter/towards-connected-governance/259735)

### Performance Investigation of Topology-Based Routing Protocols in Flying Ad-Hoc Networks Using NS-2

Sudesh Kumar and Abhishek Bansal (2020). *IoT and Cloud Computing Advancements in Vehicular Ad-Hoc Networks* (pp. 243-267).

[www.irma-international.org/chapter/performance-investigation-of-topology-based-routing-protocols-in-flying-ad-hoc-networks-using-ns-2/252296](http://www.irma-international.org/chapter/performance-investigation-of-topology-based-routing-protocols-in-flying-ad-hoc-networks-using-ns-2/252296)