LexiLogic@DravidianLangTech 2025: Political Multiclass Sentiment Analysis of Tamil X(Twitter) Comments and Sentiment Analysis in Tamil and Tulu

Billodal Roy*, Souvik Bhattacharyya*, Pranav Gupta, Niranjan Kumar M
Lowe's
Correspondence: {billodal.roy, souvik.bhattacharyya, pranav.gupta, niranjan.k.m}@lowes.com

Abstract

We present our approach and findings for two sentiment analysis shared tasks as part of DravidianLangTech@NAACL 2025. The first task involved a seven-class political sentiment classification for Tamil tweets, while the second addressed code-mixed sentiment analysis in Tamil-English and Tulu-English social media texts. We employed language-specific BERT models fine-tuned on the respective tasks, specifically utilizing the L3Cube-Tamil-BERT for Tamil classification and a Telugu-based BERT model for Tulu classification. Our system achieved notable results, particularly securing the first position in the Tulu code-mixed sentiment analysis track. The experiments demonstrate the effectiveness of language-specific pre-trained models for Dravidian language sentiment analysis, while also highlighting the challenges in handling political discourse and code-mixed content.

1 Introduction

Sentiment analysis in low-resource languages presents unique challenges, particularly for Dravidian languages with their rich morphological structure and increasing prevalence of code-mixing in social media contexts. This paper presents our unified approach to two distinct but complementary sentiment analysis tasks in the DravidianLangTech@NAACL 2025 shared task.

The first task addresses the complex challenge of political sentiment analysis in Tamil tweets, requiring fine-grained classification into seven distinct categories: Substantiated, Sarcastic, Opinionated, Positive, Negative, Neutral, and None of the above. This multi-class approach enables a more nuanced understanding of political discourse compared to traditional positive-negative sentiment classifications.

The second task focuses on sentiment analysis in code-mixed scenarios(Chakravarthi et al., 2020; Hegde et al., 2022, 2023; S. K. et al., 2024), specifically Tamil-English and Tulu-English social media comments(Durairaj et al., 2025). Code-mixing, a common phenomenon in multilingual communities, introduces additional complexity due to the interplay of linguistic features from multiple languages within the same text.

Our approach leverages recent advances in transformer-based models, specifically utilizing language-specific BERT (Devlin et al., 2019) models fine-tuned for the respective tasks. For Tamil classification, we employed L3Cube-Tamil-BERT (Joshi, 2022), while for Tulu, we innovatively adapted a Telugu-based BERT model, demonstrating the potential for cross-lingual transfer in closely related languages.[1]

2 Related Works

Sentiment analysis of textual data has been a prominent area of research for many years, with product and movie reviews being among the most extensively studied topics (Wankhade et al., 2022; S. K. et al., 2024). Many platforms include a rating system alongside text input, simplifying data preparation and framing the task as a supervised learning problem. In contrast, detecting political sentiment is considerably more challenging, as it involves distinguishing not only between political and non-political content but also identifying nuances such as sarcasm or references to specific individuals or real-time events.

Political sentiment analysis has been studied using both rule-based methods and machine learning algorithms. For instance, Elghazaly et al., 2016 utilized TF-IDF to extract document vectors and applied Support Vector Machines (SVM) and Naive

*These authors contributed equally to this work.

[1]The code for this work is available at https://github.com/prannerta100/naacl2025-dravidianlangtech

Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 635–639
May 3, 2025 ©2025 Association for Computational Linguistics

Bayes classifiers to analyze the sentiment of Arabic texts. Similarly, Bakliwal et al., 2013 demonstrated that combining a simple lexicon-based approach with bag-of-words features significantly improves accuracy. Beyond traditional supervised models like SVM and Logistic Regression, Ansari et al., 2020 employed LSTM-based (Hochreiter, 1997) models using TF-IDF features of unigrams, bigrams, and trigrams.

The rise of deep learning has transformed sentiment analysis by introducing techniques like recurrent neural networks (RNNs) and, more recently, transformers (Vaswani et al., 2017), which excel at capturing context and relationships between words. Furthermore, the development of highly parameterized large language models (LLMs) (Radford et al., 2018; Touvron et al., 2023) has made it more feasible to fine-tune models for entirely new tasks, eliminating the need to train them from scratch.

3 Dataset and Task Description

This section details the datasets and specific requirements for both the political sentiment and code-mixed sentiment analysis tasks.

3.1 Dataset Statistics

The political sentiment analysis task utilized Tamil Twitter data (Chakravarthi et al., 2025), while the code-mixed task covered Tamil-English and Tulu-English social media comments (Chakravarthi et al., 2020; Hegde et al., 2022, 2023). Tables 1, 2, and 3 present the class distributions for each dataset.

Category	Train	Test	Total
Opinionated	1,361	153	1,514
Sarcastic	790	115	905
Neutral	637	84	721
Positive	575	69	644
Substantiated	412	52	464
Negative	406	51	457
None	171	20	191
Total	4,352	544	4,896

Table 1: Political Sentiment Dataset Distribution

3.2 Task Requirements

Both tasks required sentiment classification at the message level, though with distinct objectives. The political sentiment task demanded fine-grained classification into seven categories, capturing the nuanced nature of political discourse. The classification schema included substantiated opinions, sarcasm detection, and general sentiment polarity.

Category	Train	Validation	Total
Not Tulu	4,400	543	4,943
Positive	3,769	470	4,239
Neutral	3,175	368	3,543
Mixed	1,114	143	1,257
Negative	843	118	961
Total	13,301	1,642	14,943

Table 2: Tulu-English Code-Mixed Dataset Distribution

Category	Train	Validation	Total
Positive	18,145	2,272	20,417
Unknown_state	5,164	619	5,783
Negative	4,151	480	4,631
Mixed_feelings	3,662	472	4,134
Total	31,122	3,843	34,965

Table 3: Tamil-English Code-Mixed Dataset Distribution

The code-mixed task focused on handling the complexity of bilingual text while performing sentiment classification. This task presented additional challenges due to the informal nature of social media language and the intricate patterns of language mixing. For Tamil-English, systems needed to classify texts into four categories, while Tulu-English required classification into five categories.

4 Methodology

Our approach utilized transformer-based models across all tasks, specifically leveraging language-specific BERT variants. We employed a consistent fine-tuning strategy while adapting the hyperparameters and training configurations to each task's unique requirements. Additionally, we used the text data in its raw form without any preprocessing, such as handling emojis, removing stopwords, or performing normalization.

4.1 Model Architecture

For the political sentiment analysis task, we experimented with multiple transformer based encoder and decoder models: the multilingual BERT model, a monolingual Tamil BERT model developed by L3Cube, and GPT-2 (Radford et al., 2019). The models were initialized with pre-trained weights and augmented with a classification head with a 10% dropout rate.

For the code-mixed sentiment analysis tasks, we utilized language-specific BERT models. The Tamil-English classification employed the L3Cube Tamil-BERT model, while the Tulu-English classification innovatively used the L3Cube Telugu-BERT model, leveraging the linguistic similarities between Tulu and Telugu. Each model was configured with task-specific classification heads matching their respective output dimensions: four classes for Tamil-English and five classes for Tulu-English.

4.2 Training Configuration

We implemented distinct training configurations for each task to address their specific challenges:

4.2.1 Political Sentiment Analysis

The political sentiment classifier was trained using the following configuration, as detailed in Table 4:

Parameter	Value
Learning rate	5e-5
Learning rate decay	0.9
Batch size	64
Training epochs	10
Dropout rate	10%

Table 4: Political Sentiment Training Parameters

4.2.2 Code-Mixed Sentiment Analysis

For the code-mixed tasks, we implemented separate configurations for Tamil-English and Tulu-English classification, as shown in Table 5:

Param	Tamil-En	Tulu-En
Learning rate	2e-7	2e-5
Batch Size	32	16
Epochs	5	3
Weight Decay	0.005	0.005
Label Smoothing	0.1	0.1
Grad. Acc. Steps	4	1

Table 5: Code-Mixed Training Parameters

4.3 Optimization Strategies

To address the class imbalance present in both tasks, we implemented several optimization techniques. For the code-mixed tasks, we employed label smoothing with a factor of 0.1 and weight decay of 0.005. The Tamil-English model additionally utilized gradient accumulation with 4 steps to effectively increase the batch size while managing memory constraints.

For all tasks, we utilized the AdamW optimizer and implemented mixed-precision training (FP16) to improve computational efficiency. The models were trained with early stopping based on validation loss, with checkpoints saved at each epoch. To enhance training efficiency, we employed data loading optimizations including pinned memory and multi-worker data loading for the Tamil-English task.

5 Results and Discussion

We present the results of our experiments across three sentiment analysis tasks: political sentiment analysis in Tamil and code-mixed sentiment analysis in Tamil-English and Tulu-English.

5.1 Political Sentiment Analysis

For the Tamil political sentiment classification task, our experiments with three different models showed that the Tamil-BERT model achieved the best performance with a macro F1 score of 0.36, outperforming both the multilingual BERT (0.27) and GPT-2 (0.26) models. On the final held-out test set, our system achieved a macro F1 score of 0.29, placing 9th in the competition rankings.

5.2 Code-Mixed Sentiment Analysis

The results for code-mixed tasks demonstrated notably different performance levels between Tulu-English and Tamil-English classification. Our system achieved exceptional performance on the Tulu-English task, with an overall accuracy of 0.92 and a macro F1 score of 0.84 on the validation set. This strong performance translated to the final evaluation, where our system ranked first in the competition for the Tulu-English track.

The Tamil-English task presented greater challenges, with our system achieving an accuracy of 0.45 and a macro F1 score of 0.19 on the validation set. The model showed stronger performance in identifying positive sentiments compared to other categories, but struggled with mixed feelings and unknown states.

5.3 Analysis

The disparity in performance between tasks can be attributed to several factors. The success in the Tulu-English task demonstrates the effectiveness of cross-lingual transfer learning, where a Telugu-BERT model successfully adapted to Tulu text. However, the Tamil-English task's lower performance highlights the challenges of handling larger datasets with computational constraints. The political sentiment task's moderate performance reflects the inherent complexity of fine-grained sentiment classification in political discourse.

6 Conclusion

In this paper, we presented our approach to sentiment analysis across multiple Dravidian language tasks, including political sentiment classification in Tamil and code-mixed sentiment analysis in Tamil-English and Tulu-English. Our experiments demonstrated the effectiveness of language-specific transformer models, particularly in cross-lingual scenarios, achieving first place in the Tulu-English task using a Telugu-BERT model.

The varying performance across tasks—from high accuracy in Tulu-English to moderate results in political sentiment analysis—highlights both the potential and limitations of current approaches. Our findings suggest that while transformer-based models can effectively handle complex sentiment classification tasks, their success depends significantly on factors such as dataset characteristics and the specific nature of the sentiment analysis task. Future work could focus on developing specialized architectures for political discourse analysis and improving performance on larger-scale code-mixed datasets.

7 Limitations

Our work in Dravidian language sentiment analysis, while showing promising results, has several important limitations that warrant discussion. The reliance on language-specific BERT models, while effective, introduces significant computational constraints. The models require substantial GPU resources for training, particularly evident in the Tamil-English task with its larger dataset. The need for gradient accumulation in the Tamil-English model to manage memory constraints impacts training dynamics and potentially limits model performance. A critical limitation in our approach is the absence of a dedicated pre-trained model for Tulu, unlike the availability of specific models for Tamil and Telugu. While our cross-lingual transfer from Telugu to Tulu proved successful, having a Tulu-specific pre-trained model could have potentially captured more nuanced linguistic features and improved performance further.

The datasets present their own set of challenges, exhibiting substantial class imbalances particularly evident in the political sentiment task where the "None" category comprises only 3.9% of the data. Despite implementing label smoothing and weight decay to address this imbalance, the effectiveness of our models on minority classes remains limited. The social media origin of our data introduces additional complexity - social media text often contains informal language, abbreviations, and region-specific expressions that may not be well represented in our training data. Our code-mixed sentiment analysis, focusing specifically on Tamil-English and Tulu-English combinations, may not extend well to other code-mixing patterns or to texts with more than two languages mixed together, which is common in many Indian social media contexts. For instance, a single post might contain Tamil, English, and Hindi, but our current approach cannot handle such multi-language mixing effectively.

Political sentiment analysis presents unique challenges due to its dynamic nature. Our models don't explicitly account for temporal dynamics or evolving political contexts, which is particularly important in political discourse where the meaning and sentiment of certain terms or phrases can shift rapidly based on current events. The models might need regular retraining to maintain accuracy as political discourse and language usage patterns change over time. Furthermore, political sentiment often requires understanding subtle contextual cues, historical references, and cultural nuances that may not be fully captured by our current modeling approach. These limitations point to several directions for future research, including developing more efficient architectures for code-mixed text processing, creating dedicated pre-trained models for low-resource Dravidian languages, and designing approaches that can better handle the dynamic nature of political discourse.

References

Mohd Zeeshan Ansari, Mohd-Bilal Aziz, MO Siddiqui, H Mehra, and KP Singh. 2020. Analysis of political sentiment orientations on twitter. *Procedia computer science*, 167:1821–1828.

Akshat Bakliwal, Jennifer Foster, Jennifer van der Puil, Ron O'Brien, Lamia Tounsi, and Mark Hughes. 2013. Sentiment analysis of political tweets: Towards an accurate classifier. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Vigneshwaran Muralidaran, Ruba Priyadharshini, and John Philip McCrae. 2020.

Corpus creation for sentiment analysis in code-mixed Tamil-English text. In *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, pages 202–210, Marseille, France. European Language Resources association.

Bharathi Raja Chakravarthi, Saranya Rajiakodi, Elizabeth Sherly, Thenmozhi Durairaj, Sathiyaraj Thangasamy, Ratnasingam Sakuntharaj, Prasanna Kumar Kumaresan, Kishore Kumar Ponnusamy, Arunaggiri Pandian Karunanidhi, and Rohan R. 2025. Overview of the Shared Task on Political Multiclass Sentiment Analysis of Tamil X(Twitter) Comments: DravidianLangTech@NAACL 2025. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. *Preprint*, arXiv:1810.04805.

Thenmozhi Durairaj, Bharathi Raja Chakravarthi, Asha Hedge, Hosahalli Lakshmaiah Shashirekha, Rajeswari Natarajan, Sajeetha Thavareesan, Ratnasingam Sakuntharaj, Kalyanasundaram Krishnakumari, Charumathi Rajkumar, Poorvi Shetty, and Harshitha S Kumar. 2025. Overview of the Shared Task on Sentiment Analysis in Tamil and Tulu: DravidianLangTech@NAACL 2025. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Tarek Elghazaly, Amal Mahmoud, and Hesham A Hefny. 2016. Political sentiment analysis using twitter data. In *Proceedings of the International Conference on Internet of things and Cloud Computing*, pages 1–5.

Asha Hegde, Mudoor Devadas Anusha, Sharal Coelho, Hosahalli Lakshmaiah Shashirekha, and Bharathi Raja Chakravarthi. 2022. Corpus creation for sentiment analysis in code-mixed Tulu text. In *Proceedings of the 1st Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages*, pages 33–40, Marseille, France. European Language Resources Association.

Asha Hegde, Bharathi Raja Chakravarthi, Hosahalli Lakshmaiah Shashirekha, Rahul Ponnusamy, Subalalitha Cn, Lavanya S K, Thenmozhi D., Martha Karunakar, Shreya Shreeram, and Sarah Aymen. 2023. Findings of the shared task on sentiment analysis in Tamil and Tulu code-mixed text. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, pages 64–71, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

S Hochreiter. 1997. Long short-term memory. *Neural Computation MIT-Press*.

Raviraj Joshi. 2022. L3cube-hindbert and devbert: Pre-trained bert transformer models for devanagari based hindi and marathi languages. *arXiv preprint arXiv:2211.11418*.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Lavanya S. K., Asha Hegde, Bharathi Raja Chakravarthi, Hosahalli Lakshmaiah Shashirekha, Rajeswari Natarajan, Sajeetha Thavareesan, Ratnasingam Sakuntharaj, Thenmozhi Durairaj, Prasanna Kumar Kumaresan, and Charmathi Rajkumar. 2024. Overview of Second Shared Task on Sentiment Analysis in Code-mixed Tamil and Tulu. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, Malta. European Chapter of the Association for Computational Linguistics.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models. *Preprint*, arXiv:2302.13971.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Preprint*, arXiv:1706.03762.

Mayur Wankhade, Annavarapu Chandra Sekhara Rao, and Chaitanya Kulkarni. 2022. A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55(7):5731–5780.