TamilPoliSent 2026: A Shared Task report on Multiclass Political Sentiment Analysis in Tamil

Mani Vegupatti[1], Kishore Kumar Ponnusamy[2], Bharathi Raja Chakravarthi[1], Saranya Rajiakodi[3], Durairaj Thenmozhi[4], Prasanna Kumar Kumaresan[1], Sathiyaraj Thangasamy[5]

[1]Unit for Inclusive AI, Data Science Institute, University of Galway, Ireland,
[2]Digital University Kerala, India,
[3]Central University of Tamil Nadu, India,
[4]Sri Sivasubramaniya Nadar College of Engineering, Chennai, India,
[5]Sri Krishna Adithya College of Arts and Science, India

Abstract

Political multiclass sentiment analysis involves categorizing social media comments into predefined political sentiment classes. This paper gives an overview of the TamilPoliSent 2026 shared task, which is about multiclass political sentiment analysis in Tamil. It was part of DravidianLangTech@ACL 2026. The task focused on identifying seven political sentiment categories from Tamil X (Twitter) comments in a low-resource setting. Participants were provided with annotated training and development datasets, while the test set was released without labels for final evaluation. A total of 22 teams submitted system outputs for evaluation. The submitted approaches ranged from traditional TF-IDF-based machine learning baselines to transformer fine-tuning, hybrid lexical–contextual architectures, ensemble strategies, and imbalance-aware optimization techniques. We used the macro-F1 score to check the system's performance so that all sentiment categories were fairly judged. The best-performing system achieved a macro-F1 score of 0.3935. The results highlight the challenges of modeling fine-grained political sentiment in Tamil, including class imbalance, sarcasm, informal writing styles, and semantic overlap between categories. Overall, the results show how important it is to use transformer-based models with class-balanced learning and hybrid representations for improving sentiment classification in low-resource political discourse

1 Introduction

Social media platforms such as X (formerly Twitter) have enabled the exponential rise of political comments on the World Wide Web. The ease of user participation enables real-time engagement and debates, and furthermore allows the expression of their own narratives and political sentiments. These sentiments are often multi-layered, fine-grained, and full of nuances. These characteristics necessitate sophisticated multi-class models to analyze complex political sentiments in real time and at scale. (Wankhade et al., 2022; Hermida, 2012; Velasquez, 2012; Tumasjan, 2010).

Tamil political sentiment analysis has its own challenges introduced by the nature of the language, such as rich morphology and various techniques of inflection requiring more complex processing. Moreover, the online content such as comments brings additional complexities, including code-mixing, transliteration, informal writing styles, sarcasm, implicit meanings and high noise levels making classification tasks more difficult. In addition, a low-resource setting introduces more layers of complexities, including limited annotated political data and the presence of class imbalances which impact the effectiveness of machine learning tasks. (Manoj, 2024; Shanmugavadivel, 2022; Mahata et al., 2020; Anbukkarasi and Varadhaganapathy, 2020; Thavareesan and Mahesan, 2019).

To facilitate the progress of machine learning models development in this area, the 'Political Multiclass Sentiment Analysis of Tamil X comments' shared task was introduced in 2025 (Chakravarthi et al., 2025). This task provided the participants with a 'Multi-class Tamil political sentiment analysis dataset' with seven classes - Substantiated, Sarcastic, Opinionated, Positive, Negative, Neutral, and None of the above for experimenting various techniques, building the models and benchmarking their results.

Building on the success of the 2025 shared task (Chakravarthi et al., 2025), we introduced the 2026 (Vegupatti et al., 2026) shared task to tap recent

Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 134–143
July 3, 2026

advancements in NLP over the last year and further improve the performance in the Tamil political multiclass sentiment analysis task and report the new benchmarks. The 2026 task attracted 22 participating teams, who explored a wide range of approaches, and their methodologies and details are presented in the subsequent sections.

2 Related Work

We all can agree that sentiment analysis (Tan et al., 2023; Al-Qablan et al., 2023; Wankhade et al., 2022; Birjali et al., 2021) and political sentiment analysis are among the well-researched areas within NLP supported by various classical and advanced machine learning methods, approaches and techniques. Over the years, the researchers have deployed multiple strategies for sentiment classification including lexical methods, classical supervised models, transformer-based architectures, and with the advent of Large Language Models, representation-learning-based and prompt-based approaches (Ring et al., 2024; Üveges and Ring, 2023; Tsugawa and Ohsaki, 2017; Bakliwal et al., 2013; Bermingham and Smeaton, 2011; Chung and Mustafaraj, 2011).

With the increase in Indian regional users on social media and the growing availability of Tamil sentiment datasets, research in Tamil sentiment analysis has also steadily improved over a period of years (Manoj, 2024; Priyadharshini et al., 2021; Sharmista and Ramaswami, 2020; Thavareesan and Mahesan, 2019). This growth trend is also seen in code-mixed Tamil-English sentiment analysis, where various studies and shared-tasks explored the challenges associated with hybrid-content and informal language (Shanmugavadivel, 2022; Mahata et al., 2020; Chakravarthi et al., 2020).

Whereas research in Tamil political sentiment analysis remains limited, the gap in this area led to the introduction of Multiclass Political Sentiment Analysis in Tamil shared tasks (2025–2026) (Chakravarthi et al., 2025; Vegupatti et al., 2026). These shared tasks were designed to encourage broader participation, provide curated datasets, and establish standardised benchmarks and evaluation parameters for classifying fine-grained political sentiment in a low-resource settings like Tamil.

In the 2025 (Chakravarthi et al., 2025) shared task participants experimented with a wide range of techniques, including classical machine-learning models such as SVM, Naive Bayes, ensemble methods, and feature-engineering techniques, as well as deep neural architectures like LSTM and RNN. They also explored transformer models, language-specific encoders such as IndicBERT and mBERT, and several hybrid architectures combining lexical and contextual features.

The machine learning model methodologies used by the 2026 participants are outlined in the following sections. Collectively, these efforts represent an important step toward developing NLP techniques capable of handling the linguistic diversity and political complexity found in Tamil social-media political comments and user discussions.

3 Task Description

The primary objective of this shared task is to automatically identify political discourse categories in user-generated comments collected from X (formerly Twitter). Participants are provided with annotated training and development datasets for model building and validation, while the test dataset is reserved for final evaluation. The corpus is labeled with seven predefined categories: Substantiated, Sarcastic, Opinionated, Positive, Negative, Neutral, and None of the above. Detailed task guidelines, dataset specifications, and evaluation protocols are provided on the official competition Codabench site[1].

4 Dataset

The dataset used in the TamilPoliSent 2026: Multiclass Political Sentiment Analysis in Tamil shared task consists of Tamil political comments collected from the social media platform X (formerly Twitter). The tweets that were collected show a wide range of public opinions and discussions about political leaders, parties, policies, and events in Tamil Nadu. The dataset was created to help with in-depth political sentiment analysis by categorizing tweets into seven sentiment classes: Substantiated, Sarcastic, Opinionated, Positive, Negative, Neutral, and None of the above. These categories capture different types of political discourse expressed in online discussions, enabling the analysis of nuanced sentiment patterns beyond those in traditional polarity-based sentiment analysis. We used the Python-based twscrape tool to find tweets by looking for political hashtags like #tnelection2024, #dmk, #admk, and #ntk. These hashtags are used

[1] https://www.codabench.org/competitions/11327/

to talk about major political parties and elections in Tamil Nadu. After collection, the tweets were cleaned and prepared for annotation to ensure suitability for sentiment classification research. The final curated dataset has 5,440 annotated Tamil tweets. These tweets were split into training, development, and test sets to help with model development and evaluation in the shared task. The data distribution and class distribution of training, validation, and test sets are given in Table 1.

Table 1: Data Distribution

Class	Train	Development	Test	Total
Substantiated	412	52	51	515
Sarcastic	790	115	106	1,011
Opinionated	1,361	153	171	1,685
Positive	575	69	75	719
Negative	406	51	46	503
Neutral	637	84	70	791
None of the above	171	20	25	216
Total	4,352	544	544	5,440

4.1 Annotation Process

4.1.1 Annotator Details

The dataset annotation was carried out by a team of nine annotators who had diverse academic and professional backgrounds. The annotators were undergraduate and graduate students as well as assistant professors who had previously worked with language analysis and social media discourse. The annotation team included eight male annotators and one female annotator, representing different levels of expertise and age groups. This diversity helped ensure balanced interpretation of political content and improved the reliability of the annotation process. Before the annotation phase began, all tweets were preprocessed to remove personal identifiers and sensitive information in order to protect user privacy and maintain ethical research practices when working with social media data. An example of the annotated dataset is shown in Figure 1, illustrating how tweets were labeled according to the predefined sentiment categories.

4.1.2 Annotation Guidelines and Labeling

Before the annotation process started, detailed annotation guidelines were provided to the annotators. We looked at each tweet closely and put it into one of seven predefined sentiment categories based on its content and tone. The Substantiated category includes tweets that present factual claims supported by evidence such as official announcements, statistics, or links to external sources. Tweets in the Sarcastic category use irony or humor to criticize something without being direct. Tweets in the Opinionated category share personal opinions or subjective views about political figures, policies, or events. The Positive category includes tweets that show support or appreciation toward political groups, while the Negative category includes tweets that express criticism or dissatisfaction toward political actors or decisions. The Neutral category includes tweets that convey information without clear positive or negative sentiment. The None of the above category includes tweets that do not clearly fall into any of the other defined sentiment classes. These guidelines helped annotators consistently capture nuanced expressions of political sentiment commonly found in social media discourse.

4.1.3 Inter-Annotator Agreement

To ensure the annotations were accurate and reliable, two annotators labeled each tweet separately. Only the tweets for which both annotators assigned the same sentiment label were retained in the final dataset. Tweets with conflicting annotations were removed to maintain high-quality labeled data and reduce ambiguity in the dataset. Krippendorff's alpha (α) was used to measure inter-annotator agreement. This is a common way to check reliability in multi-annotator classification tasks. This metric works well for categorical annotation settings and allows for robust estimation of agreement while accounting for chance agreement and potential disagreements between annotators.

4.1.4 Dataset Splits

After the annotation process, a total of 5,440 tweets with complete annotator agreement were retained in the final dataset. These tweets were randomly shuffled and divided into three subsets to support model training, validation, and evaluation. The training set contains 4,352 tweets, which are used by participants to train machine learning models. There are 544 tweets in the development set, which are used to tune hyperparameters and test the model. The test set also contains 544 tweets, which are used to evaluate the final system performance. The dataset shows a natural class imbalance observed in real-world political discourse on social media. For example, opinionated or sarcastic tweets are more common than neutral or evidence-based statements.

Class	Example	Translation
Substantiated	அண்ணன் இதனை சூசகமாக 11 மாதங்கள் முன்பே பேட்டியில் சொல்லிட்டார்..!	Elder brother had already said this suggestively in an interview 11 months ago..!
Sarcastic	தாய்க்காக மகள்கள், தந்தைக்காக மகன், அண்ணனுக்காக தம்பி… கணவனுக்காக மனைவி களத்தில் குடும்பங்கள்… வாரிசுகள் பராக்… பராக்… களை கட்டும் தேர்தல் பிரசாரம்	Daughters for the mother, son for the father, younger brother for the elder brother… wife for the husband—families on the trail… heirs going "parak… parak…", the election campaign in full swing.
Opinionated	இப்படிப் எதையும் எதிர்பாராமல் உழைக்கும் இளைஞர்கள் நாம் தமிழர் கட்சியில் மட்டுமே உண்டு	Only in the Naam Tamilar Party are there young people who work like this without expecting anything.
Positive	நீலகிரி மக்களவை வெற்றி வேட்பாளர் ஜெயக்குமார் அவர்களுடன் பேரறிவாளனின் தாயார் அற்புதம்மாள்	Arputham Ammal, mother of Perarivalan, with Jayakumar, the Nilgiris Lok Sabha winning candidate.
Negative	தமிழகமே உங்கள் ஆட்சியின் சீர்கேட்டை பார்த்து கறி துப்பி கொண்டிருக்கிறது; கஞ்சா போதைக்கு அடிமையாகி, தெருவுக்கு தெரு குற்றம் அதிகமானதே—தற்பொழுது நடக்கிறது உங்கள் ஆட்சியில்தான்.	Tamil Nadu, seeing the degradation of your rule, is spitting in disgust; with people enslaved to ganja intoxication, crime has increased street by street—this is what is happening now under your government.
Neutral	தென்காசி தொகுதி புதிய தமிழகம் கட்சி வேட்பாளர் டாக்டர். கிருஷ்ணசாமியை ஆதரித்து தேசிய பொதுச் செயலாளர் பிரேமலதா விஜயகாந்த் ஸ்ரீவில்லிபுத்தூரில் பிரச்சாரம்.	National General Secretary Premalatha Vijayakanth campaigned in Srivilliputhur in support of Puthiya Tamizhagam party's Tenkasi constituency candidate, Dr. Krishnasamy.
None of the above	என்ன பண்ண முடியும்?	What can be done?

Figure 1: Sample annotated comments for all classes

5 Overview of Participant Systems

A total of 22 participating teams explored various modeling paradigms for Political Multiclass Sentiment Analysis in Tamil, which range from traditional machine learning baselines to advanced transformer-based and ensemble architectures. The submitted systems can be broadly categorized into (i) classical feature-based models, (ii) transformer fine-tuning approaches, (iii) hybrid architectures combining lexical and contextual features, and (iv) ensemble and imbalance-aware frameworks.

The **Wise** team proposed a hybrid multi-view architecture for Tamil political sentiment classification. Their system combines TF-IDF features with contextual embeddings from mBERT, IndicBERT, and XLM-RoBERTa to capture both deep semantic and surface-level lexical signals. Emojis were turned into words, and short political hashtags were made longer to keep the implicit meaning. By combining multilingual and Tamil-specific representations, the system created a strong feature space for predicting sentiment in seven classes.

The **PhucNguyen** team (Nguyen and Thin, 2026) leveraged XLM-RoBERTa-base with parameter efficient fine tuning using LoRA. Low-rank adaptation was used on attention modules instead of full model fine-tuning to stop overfitting. Optuna's hyperparameter tuning found that a higher learning rate was helpful for finding minority sentiment. To find a balance between stability and sensitivity, the final submission used ensembling to combine the LoRA-adapted model with a fully fine-tuned baseline.

The **Trailblazer** team (Shanthi et al., 2026) approached the task using both classical and transformer-based methods. A TF-IDF + Linear SVM baseline was implemented alongside a fine-tuned XLM-RoBERTa model. Class-weighted training was used to mitigate imbalance, and model selection was performed using validation Macro-F1. The study highlighted the advantage of contextual embeddings in capturing subtle sentiment categories.

The **SenTamizh** team (Naren Karthik et al., 2026) adopted a hierarchical ensemble strategy. Back-translation and oversampling were applied to minority classes. IndicBERT v2 and XLM-RoBERTa models were combined with a TF-IDF SVM classifier. Additionally, specialist models for binary subtasks such as neutral vs opinionated and sarcastic vs non-sarcastic were integrated using a confidence-based gating mechanism to improve prediction reliability.

The **Codeblitz** team fine-tuned XLM-RoBERTa-base for multiclass classification and applied class-weighted cross-entropy loss to handle imbalance.

The system improved minority class performance and achieved a Macro-F1 score of 0.38 on the development set.

The **TriVector** team implemented a hybrid model combining IndicBERTv2 contextual embeddings with PCA-reduced TF-IDF features. Focal Loss with class weighting was applied during training, and logit adjustment based on class priors was used to make the model less biased toward the majority classes.

The **Semantica team** improved several multilingual encoders, such as XLM-RoBERTa, IndicBERTv2, and RemBERT, by using attention-based pooling. Partial layer unfreezing, differential learning rates, and Optuna-based hyperparameter optimization were employed. To make the final classification, model predictions were combined using a soft-voting ensemble.

The **CUET_Pinnacle** team used IndicBERTv2-MLM-only as the backbone model. The system used class-weighted loss, label smoothing, and minority-targeted augmentation through word-level perturbations. To make convergence more stable, different learning rates were used between the encoder and classifier layers.

The **Zwei Polaris** team followed a staged fine-tuning strategy using XLM-RoBERTa-base. Initially, only the classification head was trained, followed by full model fine-tuning with differential learning rates and class-weighted loss. To stop overfitting, early stopping was used based on Macro-F1.

The **TamilEcho** team (Kanimozhi Selvi et al., 2026) created a hybrid feature fusion framework that combines XLM-RoBERTa embeddings with TF-IDF n-grams, truncated SVD, emoji-based sentiment cues, and political hashtag expansion. The model was improved by using class-balanced loss and slowly unfreezing layers. It got a Macro-F1 score of 0.396.

The **PolyTicsTamil_Alchemists** team used a 5-fold cross-validation ensemble of XLM-RoBERTa-base models that were trained with Focal Loss. Back-translation and IndicBART paraphrasing were used to add to the minority classes. To make the model more general, the final predictions were made by averaging the softmax probabilities from all folds.

The **CHMOD_777** team (Arunaggiri Pandian and Prabalakshmi, 2026) used LLM-based controlled paraphrasing to fix a big class imbalance by adding more examples to the minority classes. The augmented dataset was used to fine-tune MuRIL and IndicBERT variants with Focal Loss and weighted sampling. The approach significantly improved Macro-F1 performance compared to baseline training.

The **Troopz** team developed a lightweight baseline using TF-IDF features combining word-level and character-level n-grams. A linear classifier with class-balanced training was used. The system emphasized efficiency and reproducibility while capturing Tamil morphological patterns.

The **KEC_Elevate** team fine-tuned pretrained transformer models including IndicBERT and XLM-RoBERTa Large using the HuggingFace Trainer API. Early stopping and Macro-F1-based checkpoint selection were employed to optimize performance.

The **Brain_error_call(GPT)** team used class-weighted Focal Loss to fine-tune XLM-RoBERTa by combining TF-IDF word and character n-grams with hand-crafted linguistic features. Enhanced preprocessing included emoji normalization and hashtag expansion to capture nuanced sentiment cues.

The **AstralByte** team used common machine learning methods like Support Vector Machine and Logistic Regression, which were trained on TF-IDF features. The system emphasized strong preprocessing and comparative evaluation between classifiers.

The **NamDang** team adopted a prompt-based multiclass classification framework using DSPy. Few-shot prompt optimization and typed output constraints were applied to ensure valid label generation without task-specific fine-tuning.

The **ByteBuilders** team (Mitharshana et al., 2026) constructed a weighted ensemble of XLM-RoBERTa and IndicBERT using K-fold cross-validation. To address the imbalance, they used oversampling and Focal Loss, and to represent sentences, they used mean pooling.

The **Dravidian Decoders** team (Abishek et al., 2026) fine-tuned MuRIL using Focal Loss and dynamic class weighting. Contextual data augmentation was performed for minority classes. A hybrid inference logic combined neural predictions with rule-based sarcasm detection patterns.

The **TECHACE** team fine-tuned XLM-RoBERTa for seven-class sentiment classification using class-weighted loss and GPU-accelerated training. The system outputs human-readable Tamil and English labels for practical usability.

The **PrimeLine** team fine-tuned XLM-

RoBERTa-base with a multi-task objective combining Cross-Entropy Loss and Supervised Contrastive Loss. This goal made it easier to separate closely related sentiment categories in embedding space.

The **Sentiment_kec** team focused on dataset preparation and structured annotation aligned with transformer-based learning. During preprocessing, noise was removed and the meaning was kept the same for multiclass classification.

Overall, transformer-based models dominated submissions, often enhanced with imbalance-aware objectives such as Focal Loss, class weighting, augmentation, or ensemble learning. Hybrid feature fusion methods were especially useful for dealing with sarcasm and subtle political feelings.

6 Results and Discussion

In this shared task on Political Multiclass Sentiment Analysis of Tamil X (Twitter) comments, a total of twenty-two teams submitted systems that were evaluated using Macro-averaged F1 score (MF1). Macro-F1 was chosen as the main way to measure performance to make sure that all seven sentiment categories were fairly evaluated, especially when there was an imbalance in the classes. Accuracy (ACC), Macro Precision (MP), Macro Recall (MR), Weighted Precision (WP), Weighted Recall (WR), and Weighted F1 (WF1) were also reported along with MF1.

The results show that the Wise_political team secured the first place with a Macro-F1 score of 0.3935, achieving the highest overall balanced performance across sentiment classes. The system also demonstrated strong macro precision (0.4437) and macro recall (0.3916), indicating stable discrimination across both majority and minority categories. PhucNguyen_political (Nguyen and Thin, 2026) came in second with a Macro-F1 of 0.3763, and Trailblazer_political came in third with a score of 0.3738. These top three systems show a relatively narrow performance gap, suggesting competitive modeling strategies among leading teams.

The SenTamizh_political team (Naren Karthik et al., 2026) secured fourth with a Macro-F1 of 0.3717, while Codeblitz_Tamil and TriVector_political secured fifth and sixth ranks with Macro-F1 scores of 0.3643 and 0.3626, respectively. The top six systems achieved Macro-F1 scores above 0.36, indicating comparatively better class-balanced sentiment detection capability. These systems primarily used transformer-based architectures with imbalance-aware optimization strategies.

Semantica_political (0.3589), CUET_Pinnacle_Political (0.3587), and Zwei_Polaris (0.3582) were mid-ranked systems that did well but not as well as the top-ranked systems. Their Macro-F1 scores were around 0.35–0.36. The closeness of these values indicates that several systems performed comparably, and small architectural or optimization differences influenced ranking positions.

Moderate performance was shown by systems ranked 10 to 15, such as TamilEcho_Political (Kanimozhi Selvi et al., 2026) (0.3559), PolyTicsTamil_Alchemists (0.3539), CHMOD_777_political (Arunaggiri Pandian and Prabalakshmi, 2026) (0.3425), Troopz_political (0.3374), and KEC_Elevate_political (0.3308). These results show that handling imbalances and using hybrid models helped, but separating fine-grained sentiment is still hard.

Systems that ranked lower, like Brain_error_call(GPT)_Political (0.3242), AstralByte_political (0.3091), NamDang_political (0.3004), and ByteBuilders (Mitharshana et al., 2026) (0.2987), got Macro-F1 scores below 0.33. Although these systems implemented transformer-based or baseline approaches, the results indicate difficulty in achieving balanced recall across all sentiment categories.

The final group, including DravidianDecoders_Tamil (Abishek et al., 2026) (0.2453), TECHACE (0.2134), PrimeLine (0.1744), and Sentiment_kec (0.1544), obtained comparatively lower Macro-F1 scores. The reduced performance may be attributed to challenges in class imbalance handling, limited fine-tuning strategies, or insufficient separation between closely related sentiment classes such as Opinionated and Substantiated.

The highest Macro-F1 score in this shared task was 0.3935, which shows how hard it is to classify political sentiment in Tamil social media text. The relatively moderate performance across all teams reflects several key challenges: severe class imbalance, linguistic variation, informal writing style, sarcasm usage, and semantic overlap between categories.

A clear trend observed from the results is the advantage of transformer-based models combined with imbalance-aware objectives such as Focal Loss or class-weighted cross-entropy. Systems that

Table 2: Leaderboard of Participating Systems Ranked by Macro-F1 (Political Multiclass Sentiment Analysis of Tamil X)

S.No	Team	Run	Acc	MP	MR	**Macro F1**	WP	WR	WF1	Rank
1	Wise_political	Run 1	0.4062	0.4437	0.3916	**0.3935**	0.4120	0.4062	0.3903	1
2	PhucNguyen_political	Run 3	0.3364	0.3840	0.3743	**0.3763**	0.3590	0.3364	0.3443	2
3	Trailblazer_political	Run 2	0.3474	0.3783	0.3888	**0.3738**	0.3651	0.3474	0.3454	3
4	SenTamizh_political	Run 1	0.3548	0.3831	0.3751	**0.3717**	0.3822	0.3548	0.3599	4
5	Codeblitz_Tamil	Run 1	0.3346	0.3822	0.3684	**0.3643**	0.3715	0.3346	0.3411	5
6	TriVector_political	Run 3	0.3309	0.3788	0.3642	**0.3626**	0.3734	0.3309	0.3425	6
7	Semantica_political	Run 1	0.3915	0.3840	0.3590	**0.3589**	0.3688	0.3915	0.3615	7
8	CUET_Pinnacle_Political	Run 2	0.3474	0.3720	0.3526	**0.3587**	0.3557	0.3474	0.3475	8
9	Zwei_Polaris	Run 2	0.3272	0.3914	0.3796	**0.3582**	0.3781	0.3272	0.3194	9
10	TamilEcho_Political	Run 1	0.3382	0.3666	0.3740	**0.3559**	0.3581	0.3382	0.3310	10
11	PolyTicsTamil_Alchemists	Run 1	0.3474	0.3493	0.3722	**0.3539**	0.3304	0.3474	0.3328	11
12	CHMOD_777_political	Run 1	0.3235	0.3531	0.3416	**0.3425**	0.3212	0.3235	0.3189	12
13	Troopz_political	Run 1	0.3529	0.3440	0.3331	**0.3374**	0.3447	0.3529	0.3457	13
14	KEC_Elevate_political	Run 1	0.3915	0.3409	0.3525	**0.3308**	0.3370	0.3915	0.3474	14
15	Brain_error_call(GPT)_Political	Run 1	0.3162	0.3384	0.3503	**0.3242**	0.3326	0.3162	0.2985	15
16	AstralByte_political	Run 2	0.3254	0.3065	0.3139	**0.3091**	0.3201	0.3254	0.3218	16
17	NamDang_political	Run 3	0.3346	0.3099	0.3273	**0.3004**	0.3282	0.3346	0.3183	17
18	ByteBuilders	Run 1	0.3456	0.3032	0.3075	**0.2987**	0.3031	0.3456	0.3136	18
19	Dravidian Decoders_Tamil	Run 1	0.2390	0.2530	0.3133	**0.2453**	0.1721	0.2390	0.1695	19
20	TECHACE	Run 1	0.3511	0.2468	0.2689	**0.2134**	0.2268	0.3511	0.2101	20
21	PrimeLine	Run 3	0.1875	0.2340	0.2760	**0.1744**	0.2421	0.1875	0.0916	21
22	Sentiment_kec	Run 1	0.1904	0.1603	0.1588	**0.1544**	0.1955	0.1904	0.1880	22

Acc: Accuracy; MP: Macro-Precision; MR: Macro-Recall; WP: Weighted-Precision; WR: Weighted-Recall; WF1: Weighted F1.

used both hybrid feature representations and ensemble strategies were also more stable. However, even the best-performing systems show room for improvement, indicating that fine-grained Tamil political sentiment analysis remains an open research problem.

In summary, the results demonstrate steady progress in multilingual political sentiment modeling, while also emphasizing the need for improved class-balanced learning, culturally aware representation modeling, and advanced data augmentation strategies for future work.

7 Conclusion

The TamilPoliSent 2026: Multiclass Political Sentiment Analysis in Tamil shared task offered important insights into fine-grained political discourse classification for a low-resource Dravidian language widely used on social media. Conducted as part of DravidianLangTech@ACL 2026, the task challenged twenty-two participating teams to classify Tamil X (Twitter) comments into seven sentiment categories. They used a variety of methods, from traditional TF-IDF baselines to transformer-based fine-tuning, hybrid feature fusion, ensemble strategies, and imbalance-aware optimization techniques. System performance was evaluated using Macro-F1 to ensure balanced assessment across sentiment classes, with the best system achieving a Macro-F1 score of 0.3935 and several competitive systems performing within the 0.35–0.37 range. The results highlight the difficulty of modeling sarcasm, semantic overlap between categories, class imbalance, and code-mixed informal language in Tamil political discourse. The shared task shows that transformer-based architectures perform better when they are combined with class-balanced learning goals and hybrid lexical–contextual representations. However, there is still a lot of room for improvement. Future work may benefit from advanced data augmentation, culturally grounded representation learning, explainable AI techniques, and more robust imbalance reduction strategies to improve fairness, interpretability, and reliability in political sentiment analysis for Tamil and other low-resource languages.

8 Limitations

Although the shared task provides a useful benchmark for multiclass political sentiment analysis in Tamil, several limitations remain. One limitation is that the data comes from X (formerly Twitter), which could lead to biases that are specific to that platform and an uneven representation of political views. There are often informal writing styles, spelling mistakes, emojis, and code-mixed

Table 3: Summary of Participating Systems for Political Multiclass Sentiment Analysis (✓= reported, ✗= not reported in the team description).

Team	Preproc	Ind/PT	Ens	TF-IDF	Aug	Imb	Thr	LLM/LoRA	Rule
Wise	✓	✓	✓	✓	✗	✗	✗	✗	✗
PhucNguyen (Nguyen and Thin, 2026)	✗	✓	✓	✗	✗	✓	✓	✓	✗
Trailblazer (Shanthi et al., 2026)	✗	✓	✗	✓	✗	✓	✗	✗	✗
SenTamizh (Naren Karthik et al., 2026)	✗	✓	✓	✓	✓	✓	✓	✗	✗
Codeblitz	✗	✓	✗	✗	✗	✓	✗	✗	✗
TriVector	✗	✓	✗	✓	✗	✓	✓	✗	✗
Semantica	✗	✓	✓	✗	✗	✓	✓	✗	✗
CUET_Pinnacle	✓	✓	✗	✗	✓	✓	✗	✗	✗
Zwei_Polaris	✓	✓	✗	✗	✗	✓	✓	✗	✗
TamilEcho (Kanimozhi Selvi et al., 2026)	✓	✓	✗	✓	✗	✓	✗	✗	✗
PolyTicsTamil_Alchemists	✗	✓	✓	✗	✓	✓	✗	✗	✗
CHMOD_777	✗	✓	✗	✗	✓	✓	✗	✓	✗
Troopz	✗	✗	✗	✓	✗	✓	✗	✗	✗
KEC_Elevate	✓	✓	✗	✗	✗	✗	✓	✗	✗
Brain_error_call	✓	✓	✗	✓	✗	✓	✗	✗	✗
AstralByte	✓	✗	✗	✓	✗	✗	✗	✗	✗
NamDang	✓	✗	✗	✗	✗	✗	✗	✓	✗
ByteBuilders (Mitharshana et al., 2026)	✓	✓	✓	✗	✓	✓	✗	✗	✗
Dravidian_Decoders (Abishek et al., 2026)	✗	✓	✗	✗	✓	✓	✗	✗	✓
TECHACE	✗	✓	✗	✗	✗	✓	✗	✗	✗
PrimeLine	✗	✓	✗	✗	✗	✓	✓	✗	✗
Sentiment_kec	✓	✗	✗	✗	✗	✗	✗	✗	✗

Preproc: Explicit preprocessing (emoji/hashtag normalization); **Ind/PT**: Indic or multilingual pretrained transformer; **Ens**: Ensemble strategy; **TF-IDF**: Lexical feature integration; **Aug**: Data augmentation; **Imb**: Class imbalance handling; **Thr**: Threshold tuning/calibration; **LLM/LoRA**: Prompting or LoRA fine-tuning; **Rule**: Rule-based or hybrid symbolic component.

language in social media posts, which can make it harder for computers to figure out how people feel. Regional dialect variations and frequent mixing of Tamil with English further increase linguistic complexity and affect model performance. Another limitation arises from the class imbalance present in the dataset, where certain categories such as opinionated or sarcastic comments appear more frequently than others. This imbalance can influence model learning and reduce the ability of systems to correctly identify minority sentiment categories. In addition, the subtle boundaries between sentiment classes such as opinionated, substantiated, and sarcastic make the classification task inherently difficult, even for human annotators. The dataset size is also relatively limited compared to large-scale sentiment corpora, which restricts the ability of deep learning models to fully capture the diversity of political expressions in Tamil. These limitations highlight the need for larger, more balanced datasets, improved handling of code-mixed and

Ethical Considerations

This shared task dataset contains Tamil political comments collected from social media platforms. They may include strong opinions, criticism, sarcasm, or even rude political language expressed by users. The content is provided strictly for research and educational purposes and does not reflect the views of the authors or organizers. All tweets were collected from publicly available sources, and efforts were made to remove direct personal identifiers wherever possible to protect user privacy. Only the minimal information required for the shared task was released. The annotation process was carried out by trained Tamil-proficient annotators who followed clear guidelines during labeling. Although the dataset is intended to support research in political sentiment analysis and responsible content understanding, it may still contain biases inherent in social media discourse and political discussions. Therefore, the dataset should be used carefully and responsibly, with attention to potential risks such as bias amplification, misinterpretation of political opinions, or misuse in automated decision-making systems. We encourage transparent research practices and recommend human oversight when deploying models trained on this dataset in real-world applications.

Acknowledgments

The authors, Bharathi Raja Chakravarthi was funded by a research grant from Research Ireland under grant number SFI/12/RC/2289_P2 (Insight_2) and Prasanna Kumar Kumaresan funded by Research Ireland Centre for Research Training in Artificial Intelligence under Grant No. 18/CRT/6223.

References

Ramaswami Abishek, Rohith P P, Kogilavani Shanmugavadivel, Gothai E, and Malliga Subramanian. 2026. Dravidian Decoders@DravidianLangTech 2026: Political Multiclass Sentiment Analysis of Tamil X (Twitter) via Asymmetric Weighted Cross-Entropy. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Tamara Amjad Al-Qablan, Mohd Halim Mohd Noor, M. Al-Betar, and A. Khader. 2023. A survey on sentiment analysis and its applications. *Neural Computing and Applications*, 35:21567–21601.

S. Anbukkarasi and S. Varadhaganapathy. 2020. Analyzing sentiment in tamil tweets using deep neural network. *ICCMC*, pages 449–453.

Karunanidhi Arunaggiri Pandian and Arumugam Prabalakshmi. 2026. CHMOD_777@DravidianLangTech 2026: LLM-Augmented Transformer Fine-Tuning for Tamil Political Sentiment Analysis. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

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. In *Proceedings of the Workshop on Language Analysis in Social Media*, pages 49–58, Atlanta, Georgia. Association for Computational Linguistics.

Adam Bermingham and Alan Smeaton. 2011. On using Twitter to monitor political sentiment and predict election results. In *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011)*, pages 2–10, Chiang Mai, Thailand. Asian Federation of Natural Language Processing.

Marouane Birjali, Mohammed Kasri, and A. B. Hssane. 2021. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowl. Based Syst.*, 226:107134.

Bharathi Raja Chakravarthi, R. Priyadharshini, Vigneshwaran Muralidaran, Shardul Suryawanshi, Navya Jose, E. Sherly, and John P. McCrae. 2020. Overview of the track on sentiment analysis for dravidian languages in code-mixed text. *Proceedings of the 12th Annual Meeting of the Forum for Information Retrieval Evaluation*.

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 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*, pages 104–111.

Jessica Chung and Eni Mustafaraj. 2011. Can collective sentiment expressed on twitter predict political elections? In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 25, pages 1770–1771.

Alfred et al. Hermida. 2012. Share, like, recommend: Decoding the social media news consumer. *Journalism Studies*.

CS Kanimozhi Selvi, NS Inigashree, AG Moneissh, and J Kavinraj. 2026. TamilEcho@DravidianLangTech 2026: Hybrid XLM-RoBERTa with Sarcasm-Aware Feature Fusion for Political Multiclass Sentiment Analysis in Tamil X. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Sainik Kumar Mahata, Dipankar Das, and Sivaji Bandyopadhyay. 2020. Sentiment classification of code-mixed tamil-english tweets using birnn. *arXiv preprint arXiv:2010.10111*.

S. Manoj. 2024. A survey on sentiment analysis in tamil: Critical analysis. In *Annals of Computer Science and Information Systems*, volume 42, pages 49–62.

TV Mitharshana, S Shanthi, V Lavana, and R Kaviya Varma. 2026. Byte-Builders@DravidianLangTech 2026: Transformer-Based Weighted Ensemble for Political Multiclass Sentiment Analysis of Tamil X (Twitter) Comments. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Kandasamy Naren Karthik, Aakash Balasubramanian, Hemnath Dillibabu, Aadhithya Srinivasan Rajashree, and Vijaya Lakshmi A. 2026. Sen-Tamizh@DravidianLangTech 2026: A Specialist-Augmented Transformer Ensemble for Multiclass Political Sentiment Analysis of Tamil X Comments. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Dinh Khac Phuc Nguyen and Dang Van Thin. 2026. PhucNguyen@DravidianLangTech 2026: Political Multiclass Sentiment Analysis with XLM-RoBERTa and Low-Rank Adaptation. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Ruba Priyadharshini, Bharathi Raja Chakravarthi, Sajeetha Thavareesan, Dhivya Chinnappa, Durairaj Thenmozhi, and Rahul Ponnusamy. 2021. Overview of the dravidiancodemix 2021 shared task on sentiment detection in tamil, malayalam, and kannada. In *Proceedings of the 13th Annual Meeting of the Forum for Information Retrieval Evaluation*, pages 4–6.

Orsolya Ring, Martina Katalin Szabó, Csenge Guba, Bendegúz Váradi, and István Üveges. 2024. Approaches to sentiment analysis of hungarian political news at the sentence level. *Language Resources and Evaluation*, 58(4):1233–1261.

Kogilavani et al. Shanmugavadivel. 2022. An analysis of machine learning models for sentiment analysis of tamil code-mixed data. *Computer Speech & Language*, 76:101407.

Murugan Shanthi, Anbuaruvi R, and Anuradha C. 2026. TrailBlazer@DravidianLangTech 2026: A Comparative Study of TF-IDF SVM and XLM-RoBERTa for Political Multiclass Text Classification. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

A. Sharmista and M. Ramaswami. 2020. Sentiment analysis on tamil reviews as products in social media using machine learning techniques: A novel study. Madurai Kamaraj University, Madurai-625 021. Technical Report.

Kian Long Tan, C. Lee, and K. Lim. 2023. A survey of sentiment analysis: Approaches, datasets, and future research. *Applied Sciences*.

Sajeetha Thavareesan and Sinnathamby Mahesan. 2019. Sentiment analysis in tamil texts: A study on machine learning techniques. In *Proceedings of ICIIS*, pages 320–325.

Sho Tsugawa and Hiroyuki Ohsaki. 2017. On the relation between message sentiment and its virality on social media. *Social network analysis and mining*, 7:1–14.

Andranik et al. Tumasjan. 2010. Predicting elections with twitter. In *AAAI ICWSM*, pages 178–185.

István Üveges and Orsolya Ring. 2023. Hunembert: a fine-tuned bert-model for classifying sentiment and emotion in political communication. *IEEE Access*, 11:60267–60278.

Mani Vegupatti, Kishore Kumar Ponnusamy, Bharathi Raja Chakravarthi, Saranya Rajiakodi, Durairaj Thenmozhi, Prasanna Kumar Kumaresan, and Sathiyaraj Thangasamy. 2026. TamilPoliSent 2026: A Shared Task report on Multiclass Political Sentiment Analysis in Tamil. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Alcides Velasquez. 2012. Social media and online political discussion. *New Media & Society*.

M. Wankhade, A. C. S. Rao, and C. Kulkarni. 2022. A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55(8):5731–5780.