Overview of the Shared Task on Sentiment Analysis in Tamil and Tulu

Durairaj Thenmozhi[1], **Bharathi Raja Chakravarthi**[2], **Asha Hedge**[3], **Hosahalli Lakshmaiah Shashirekha**[3], **Rajeswari Natarajan**[4], **Sajeetha Thavareesan**[5], **Ratnasingam Sakuntharaj**[5] **Krishnakumari Kalyanasundaram**[6] **Charumathi Rajkumar**[7] **Poorvi Shetty**[8] **Harshitha S Kumar**[3]

[1]Sri Sivasubramaniya Nadar College of Engineering, Tamil Nadu, India,
[2]School of Computer Science, University of Galway, Ireland,
[3]Mangalore University, Mangalore, India,
[4]SASTRA University, SRC campus, Kumbakonam, Tamil Nadu, India,
[5]Eastern University, Sri Lanka, [6]A.V.C. College of Engineering, Tamil Nadu, India
[7]The American College, Madurai, Tamil Nadu, India, [8]JSS College, Mysore, India.

Abstract

Sentiment analysis is an essential task for interpreting subjective opinions and emotions in textual data, with significant implications across commercial and societal applications. This paper provides an overview of the shared task on Sentiment Analysis in Tamil and Tulu, organized as part of DravidianLangTech@NAACL 2025. The task comprises two components: one addressing Tamil and the other focusing on Tulu, both designed as multi-class classification challenges, wherein the sentiment of a given text must be categorized as positive, negative, neutral and unknown. The dataset was diligently organized by aggregating user-generated content from social media platforms such as YouTube and Twitter, ensuring linguistic diversity and real-world applicability. Participants applied a variety of computational approaches, ranging from classical machine learning algorithms such as Traditional Machine Learning Models, Deep Learning Models, Pre-trained Language Models and other Feature Representation Techniques to tackle the challenges posed by linguistic code-mixing, orthographic variations, and resource scarcity in these low resource languages.

1 Introduction

Sentiment Analysis (SA), a computational approach to deciphering human opinions and emotions through written language, has become increasingly crucial across various domains such as social media monitoring, market research, and customer feedback analysis (Wankhade et al., 2022). The digital communication landscape has witnessed a growing linguistic phenomenon known as code-mixing, which is particularly prevalent in multilingual societies (Alam et al., 2024). Code-mixing represents the intricate practice of interweaving multiple languages within a single communicative context, a trend that has garnered significant attention in Natural Language Processing, especially in linguistically diverse regions like India (Sampath and Supriya, 2024). An enhanced sentiment dictionary that incorporates both labeled and unlabeled data from source and target domains significantly improves sentiment classification in multi-domain contexts (Sivasankar et al., 2021). This approach is not only applicable to domain adaptation within the same language but can also be extended to sentiment classification across different languages. By leveraging such cross-lingual adaptation, models can better handle language-specific nuances, improving sentiment analysis in low-resource languages like Tamil and Tulu, as well as facilitating sentiment classification between languages with distinct linguistic features. Systems trained on monolingual data face challenges with code-mixed text because of the intricate nature of code-switching across different linguistic levels. (Ponnusamy et al., 2023). The complexity of SA escalates when confronting code-mixed text, as traditional analysis methods struggle with the nuanced linguistic variations introduced by script and language mixing (Perera and Caldera, 2024). Users frequently leverage Latin script and common English words, creating hybrid textual landscapes that challenge conventional sentiment extraction techniques (Hegde et al., 2022). The intricate nature of code-mixing further compounds sentiment analysis challenges, with individuals often switching between scripts and languages in unpredictable ways (Chakravarthi et al., 2021; Sambath Kumar et al., 2024). These linguistic complexities highlight the critical need for more advanced analytical

Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 97–103
May 3, 2025 ©2025 Association for Computational Linguistics

approaches. Addressing these code-mixed linguistic complexities is important to capture modern communication's rich, dynamic nature accurately.

Granted classical language status by the Indian government in 2004, Tamil boasting a literary heritage that spans over two millennia and is one of the world's most enduring classical languages (Abirami et al., 2024). Beyond its status as the official language of Tamil Nadu and Puducherry, the language has transcended geographical boundaries, finding vibrant expression in diverse global communities including Malaysia, Mauritius, Fiji, and South Africa (Rajalakshmi et al., 2023).

Tulu, a member of the Dravidian language family, boasts over three million speakers known as Tuluvas, primarily concentrated in Karnataka's Dakshina Kannada and Udupi districts, with additional communities extending to Mumbai and Gulf countries (Hegde et al., 2022). The language has carved out a significant digital footprint, with active engagement across social media platforms and a thriving film industry that further amplifies its cultural relevance (Narayanan and Aepli, 2024).

This shared task presents a new corpus in the Tamil and Tulu languages. We used comments and posts of Movie reviews from Youtube for this shared task of Sentiment Analysis.

2 Task Description

The goal of this shared task[1] is to identify the sentiment polarity of the code-mixed dataset of comments or posts in Tamil-English and Tulu-English collected from social media. The comment or post may contain more than one sentence but the average sentence length of the corpora is one. Each comment or post is annotated with sentiment polarity at the comment or post level. These code-mixed datasets consist of posts and comments collected from YouTube comments. Our proposal aims to encourage research that will reveal how sentiment is expressed in code-mixed scenarios on social media. For every comment in Tamil and Tulu, the objective is to classify it into positive, negative, neutral, or mixed emotions.

3 Dataset Description

Recent advancements in natural language processing (NLP), particularly transformer-based models and multilingual embeddings, have further accelerated research in sentiment analysis. Additionally, the integration of large language models (LLMs) and zero-shot learning techniques has improved sentiment classification accuracy for underrepresented languages, enabling better contextual understanding and real-time analysis. For the analysis of sentiment in YouTube comments, two meticulously curated datasets—i) Tamil-English and ii) Tulu-English—have been introduced to support computational linguistics research. These resources act as essential linguistic benchmarks, aiding in the exploration of hybrid-language processing. By offering diverse and naturally occurring text samples, they assist scholars and industry professionals in enhancing machine learning models tailored for multilingual and phonetically transcribed content. Additionally, these datasets play a pivotal role in refining Artificial Intelligence systems to better interpret the emotional tone and contextual intricacies of underrepresented languages in digital discourse. The dataset is prepared in two languages such as Tamil and Tulu as listed in Table 1.

3.1 Tamil Data

We gathered comments and posts from YouTube related to various Tamil films, encompassing discussions, reviews, and audience opinions. This data includes user perspectives on different aspects such as storyline, performances, music, and overall cinematic experience. A sample set of comments are listed in the Table 2.

3.2 Tulu Data

We collected comments and posts from YouTube about various Tulu films, covering discussions, reviews, and audience opinions. The data reflects user perspectives on elements such as storyline, performances, music, and overall cinematic appeal. A sample set of comments are listed in the Table 3.

4 Methodologies used in the Submission

Team Hermes fine-tuned the pre-trained multilingual transformer model, cardiffnlp/twitter-xlm-roberta-base-sentiment from Hugging Face. A PyTorch-based data pipeline with a custom dataset class and DataLoaders was used to handle batched input. The model employed AdamW optimization, and early stopping based on validation F1 score.

Team byteSizedLLM (Manukonda and Kodali, 2025) used hybrid approach combined a fine-tuned XLM-RoBERTa base model with a customized attention BiLSTM network to leverage contextualized embeddings and sequential modeling. The

[1] https://codalab.lisn.upsaclay.fr/competitions/20893

Label	Train Set		Development Set		Test Set	
	Tamil	**Tulu**	**Tamil**	**Tulu**	**Tamil**	**Tulu**
Positive	18,145	3,769	2,272	470	1,983	453
Negative	4,151	843	480	118	458	88
Neutral	5,164	3,175	619	368	593	343
Mixed	3,662	1,114	472	143	425	120

Table 1: Distribution of data in Train, Development, and Test sets for Tamil and Tulu languages

S.No	Text	Label
1	Therikaaa vidalamaanu kealvi yellam keadayadhu... Iranginaale theri dhaaaa	Positive
2	Romba naalaki aprama suriya annana ipdi pakuravanga mattum solluga	Unknown_state
3	Aiooo..samy mudilada..yenda 2D ku inoru flop conform	Negative
4	The word vera level thalaiva unaku vayase agadha paaaa	Mixed_feelings

Table 2: Sample set of Tamil Comments with Labels

XLM-RoBERTa model was fine-tuned using MLM on a small portion of the AI4Bharat dataset, enriched with fully and partially transliterated text to improve multilingual and transliteration handling. Attention and BiLSTM layers were used to enhance sequential dependency capture.

For the Tamil dataset, XNet, Naive Bayes, and Logistic Regression were employed by Team RMK-Mavericks to predict sentiment, leveraging their ability to capture diverse text patterns and nuances. For the Tulu dataset, they used SVM, Random Forest, and Logistic Regression. TF-IDF vectorization transformed text into numerical features, with hyperparameter tuning to optimize results.

Team codecrackers (P et al., 2025) system employs three models—Naive Bayes, SVM, and an LSTM-based deep learning model—to address sentiment analysis in Tamil code-mixed text. Naive Bayes and SVM leverage TF-IDF vectorization and traditional machine learning techniques for simpler patterns, while the LSTM-based model combines word-level Word2Vec embeddings and character-level features to capture complex syntactic and semantic nuances.

(Sreeja and Bharathi, 2025) fine-tuned a pre-trained transformer model, distilroberta-base, for multilingual sentiment analysis. Their program preprocesses data by cleaning text, mapping labels to numeric values, and tokenizing inputs, while addressing class imbalances by incorporating calculated class weights into the loss function. Optimized training techniques like gradient accumulation and mixed precision enhanced efficiency.

Team Code Conquerors employed data preprocessing that involved addressing class imbalance using the class-weight method, and resolving out-of-vocabulary issues by developing a vocabulary. The hybrid model begins with an embedding layer, followed by a Conv1D layer and a max pooling layer. A BiLSTM layer was used, with fully connected dense layers and a softmax layer for finding the sentiments.

Team ET2025 used mBERT model (Adyanthaya, 2025) fine-tuned for sentiment classification in Tamil and Tulu. Preprocessed datasets were tokenized and split into training and evaluation sets, with the Hugging Face Trainer API used for training and evaluation.

Team CIC used Feature extraction with Logistic Regression which involved identifying the top bigrams and trigrams based on frequency to capture class similarities effectively. These enabled the model to distinguish between classes more accurately.

Team SKV Trio combined TF-IDF and BERT embeddings. TF-IDF embeddings were reduced to 512 dimensions via an RBF Sampler for efficiency, while the bert-base-uncased model generated con-

S.No	Text	Label
1	Edde msg koryar prasamsha	Positive
2	Enchi pankda comedy	Negative
3	Kas ejjande boys yerla hotel popujer. Ponnulu mathra	Mixed
4	Padhyana ganapathi bhagavathike	Neutral
5	Well done Keep It up!!!	Not Tulu

Table 3: Sample set of Tulu Comments with Labels

textual embeddings by averaging token representations. Random Forest classifier was trained on the merged features. Model generalizability was verified through 5-fold cross-validation.

Pre-processing techniques such as tokenization, special character removal, and stopword filtering were applied by Team YenLP_CS. TF-IDF was used for feature extraction (Adyanthaya, 2025), followed by training an ensemble of Random Forest (fine-tuned with GridSearchCV) and SVM (Shanmugavadivel et al., 2025). Word2Vec embeddings were generated, supporting LSTM and BiLSTM deep learning models. Finally, the multilingual transformer model mBERT was fine-tuned on the task data.

Team KECTechTitans used three models - KNN, SVM, and Decision Tree on data vectorized by TF-IDF. KNN was selected for its simplicity in proximity-based classification, SVM for its effectiveness in high-dimensional data, and Decision Tree for its interpretability. Model performance compared to determine the most effective approach.

Team Dynamic_Crew cleaned text data and normalized it to remove noise, such as special characters and stop words, while addressing language-specific nuances. For feature extraction, Count Vectorizer and TF-IDF Vectorizer were used. Classifiers like Decision Tree, Random Forest, and KNN were employed by this team.

Team Team_Mavericks used TF-IDF Vectorizer to capture term importance with unigrams and bigrams. Data preprocessing included combining training and validation datasets, splitting for evaluation, and vectorizing the text. The Random Forest model was used, optimized using GridSearchCV for hyperparameter tuning.

Team JustATalentedTeam (Ponsubash Raj R, 2025) began with transliterating texts to the English script when necessary. Two methodologies were used: 1) a Logistic Regression model with TF-IDF Vectorizer using a character-level analyzer, and 2) a combined approach involving tokenization, FastText embeddings, and a deep learning model for sentiment classification of code-mixed text.

Team lemlem used google-bert/bert-base-multilingual-uncased model, with a classification head consisting of a dense layer and softmax activation to predict sentiment categories. By fine-tuning the model, the system aimed to subtle syntactic and semantic nuances, performing sentiment analysis even with limited annotated data.

Team MysticCIOL's approach involved using custom pre-trained models, each specifically trained on general Tamil and Tulu data. A Multi-Layer Perceptron was applied on top of the pre-trained embeddings to fine-tune them for sentiment classification. The fine-tuned models were then used to generate predictions on the test data.

In Team Cognitext's approach, the text data was preprocessed by converting it to lowercase, removing URLs, mentions, hashtags, and special characters. TF-IDF vectorization was applied to extract feature through unigrams and bigrams. A Logistic Regression classifier was then trained on it. The model's performance was evaluated using a validation set.

Team TensorTalk (Anishka and J, 2025) used a combination of SVM, Logistic Regression, and Random Forest classifiers. The preprocessing pipeline involved cleaning the text, removing stopwords, performing lemmatization, and applying TF-IDF vectorization. To address the class imbalance in the datasets, we used the Synthetic Minority Oversampling Technique (SMOTE) to generate synthetic samples for underrepresented classes.

Team SSNTrio (J et al., 2025) explored the use of Multilingual BERT and language-specific Tamil

BERT models. To address the class imbalance, random upsampling was applied. The text data was tokenized using the appropriate tokenizer for each model. The models were fine-tuned using the training set, with optimized hyperparamters.

Team Anna-CIOL used custom pre-trained models specifically fine-tuned for Tamil and Tulu to extract embeddings. They employed a Multi-Layer Perceptron to fine-tune these embeddings for sentiment classification. Once the models were fine-tuned, they used them to generate predictions.

Team lowes fine-tuned a language-specific BERT model pre-trained by l3cube-pune using the provided datasets, implementing class weighting to handle imbalanced labels and optimizing training parameters. The model handled class imbalance through oversampling and by using a weighted model to compute loss during training.

The preprocessing techniques, feature extractions, and classifiers used by the participating teams are summarized below.

4.1 Preprocessing

Teams have used preprocessing techniques namely stop word removal, removal of hashtags & URLS and lemmatization in their approaches. They have used random upsampling and SMOTE methods for handling data imbalance problems.

4.2 Feature Extraction

Submitting teams used unigrams, bigrams, trigrams and Character-level features, and vectorized the text with TF-IDF, Word2Vec, fasttext and BERT embeddings.

4.3 Classifiers

Participants used traditional classifiers namely, SVM, logistic regression, multilayer perceptron, random forest (Gowda, 2025), decision tree, KNN and Naive Bayes approaches for finding the sentiments. Deep learning frameworks namely LSTM and BILSTM are used by the participants (Srichandra et al., 2025; Rajalakshmi et al., 2023). Transformer models namely distilRoberta, Multilingual BERT and XLM-Roberta are used by the participants among which XLM-Roberta performs better when compared to other approaches (Krasitskii et al., 2025) (G et al., 2025). Further, language specific BERT pretrained by l3cube-pune performs better for Tulu language with 0.5938 as F1-score. A pretrained model finetuned on AI4Bharart dataset performs better for Tamil language with an F1-score of 0.5036.

5 Results

The participating teams submitted 2 to 3 runs to the both Tamil and Tulu tasks. 21 teams participated in Tamil task and 20 teams participated in Tulu task. Their submissions were evaluated and ranked based on macro F1-score. Scores are tabulated in Tables 4 and 5 for Tamil and Tulu sub tasks respectively.

Team Name	Macro F1 Score	Rank
byteSizedLLM	0.5036	1
ET2025	0.4986	2
Hermes	0.4957	3
JustATalentedTeam	0.4919	4
Lemlem	0.4709	5
SSNTrio	0.4461	6
CIC	0.4409	7
codecrackers	0.4389	8
KECTechTitans	0.4386	9
RMKMavericks	0.4354	10
MysticCIOL	0.4299	11
KEC-Elite-Analysts	0.4131	12
YenLP_CS	0.4117	13
Team_Mavericks	0.4011	14
Dynamic_crew	0.3852	15
lowes	0.3834	16
SSN_IT_SENTI	0.3799	17
CodeConquerors	0.3357	18
Anna-CIOL	0.335	19
Cognitext	0.2867	20
TensorTalk	0.2427	21

Table 4: Rank list for Tamil sentiment analysis.

6 Conclusion

There are 21 teams participated in the Tamil subtask and 20 teams participated in the Tulu subtask. Participants used preprocessing techniqes like lemmatization, removal of stop word, URLs and hash tags in their approaches. Teams who have employed traditional classifiers used n-gram features with TF-IDF scores and static embeddings namely Word2Vec and FastText for vectorization. Most of the teams used transformer models namely multilingual BERT, distilRoberta and XLM-Roberta, Pretrained models finetuned on AI4Bharat dataset are used by the teams. Team 'byteSizedLLM' (Manukonda and Kodali, 2025) who used language

Team Name	Macro F1 Score	Rank
lowes	0.5938	1
ET2025	0.5882	2
Hermes	0.5801	3
JustATalentedTeam	0.5617	4
SSNTrio	0.5609	5
Lemlem	0.5583	6
YenLP_CS	0.5511	7
codecrackers	0.5425	8
RMKMavericks	0.5318	9
TensorTalk	0.5269	10
Team_Mavericks	0.4683	11
CIC	0.4509	12
CodeConquerors	0.4357	13
SSN_IT_SENTI	0.3904	14
Anna-CIOL	0.3863	15
SKV-trio	0.3767	16
Dynamic_crew	0.3750	17
KECTechTitans	0.3197	18
MysticCIOL	0.1546	19
Cognitext	0.1491	20

Table 5: Rank list for Tulu sentiment analysis.

BERT pretrained on AI4Bharat dataset secure first position in the Tamil subtask, and the team "lowes" who used pretrained models created by l3cube-pune secured first position in the Tulu subtask.

Acknowledgments

This work was conducted with the financial support from Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289_P2(Insight_2), supported in part of Science Foundation Ireland Centre for Research Training in Artificial Intelligence under Grant No. 18/CRT/6223.

References

A M Abirami, Wei Qi Leong, Hamsawardhini Rengarajan, D Anitha, R Suganya, Himanshu Singh, Kengatharaiyer Sarveswaran, William Chandra Tjhi, and Rajiv Ratn Shah. 2024. Aalamaram: A large-scale linguistically annotated treebank for the Tamil language. In *Proceedings of the 7th Workshop on Indian Language Data: Resources and Evaluation*, pages 73–83, Torino, Italia. ELRA and ICCL.

Raksha Adyanthaya. 2025. Sentiment analysis on code-mixed tamil and tulu data using machine learning and deep learning models. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Sadia Alam, Md Farhan Ishmam, Navid Hasin Alvee, Md Shahnewaz Siddique, Md Azam Hossain, and Abu Raihan Mostofa Kamal. 2024. BnSentMix: A diverse Bengali-English code-mixed dataset for sentiment analysis.

K Anishka and Anne Jacika J. 2025. Sentiment analysis in tamil and tulu-dravidianlangtech@naacl2025. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Prasanna Kumar Kumaresan, Ratnasingam Sakuntharaj, Anand Kumar Madasamy, Sajeetha Thavareesan, Bhavukam Premjith, K R Sreelakshmi, Subalalitha Chinnaudayar Navaneethakrishnan, John, Patrick McCrae, and Thomas Mandl. 2021. Overview of the hasoc-dravidiancodemix shared task on offensive language detection in tamil and malayalam. In *In Proceedings of the Forum for Information Retrieval and Evaluations*.

Jyothish Lal G, Premjith B, Bharathi Raja Chakravarthi, Saranya Rajiakodi, Bharathi B, Rajeswari Natarajan, and Ratnavel Rajalakshmi. 2025. Overview of the shared task on multimodal hate speech detection in dravidian languages: Dravidianlangtech@naacl 2025. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Anusha M D Gowda. 2025. Bridging linguistic complexity: Sentiment analysis of tamil code-mixed text using meta-model. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

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.

Bhuvana J, Mirnalinee T T, Diya Seshan, Rohan R, and Avaneesh Koushik. 2025. Ssntrio@dravidianlangtech 2025: Sentiment analysis in dravidian languages using multilingual bert. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Mikhail Krasitskii, Olga Kolesnikova, Grigori Sidorov, and Alexander Gelbukh. 2025. Multilingual sentiment analysis: Understanding tamil-english code-mixing with transformer models. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language*

Technologies for Dravidian Languages. Association for Computational Linguistics.

Durga Prasad Manukonda and Rohith Gowtham Kodali. 2025. Sentiment analysis in tamil using transliteration-aware xlm-roberta and attention-bilstm. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Manu Narayanan and Noëmi Aepli. 2024. A Tulu resource for machine translation. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 1756–1767, Torino, Italia. ELRA and ICCL.

Lalith Kishore V P, Dr. G Manikandan, Mohan Raj M A, Keerthi Vasan A, and Aravindh M. 2025. odecrackers@dravidianlangtech 2025: Sentiment classification in tamil and tulu code-mixed social media text using machine learning. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Perera and Caldera. 2024. Sentiment analysis of code-mixed text: A comprehensive review. *Int. J. Comput. Sci. Netw. Secur.*, 24(11):73–84.

Kishore Kumar Ponnusamy, Charmathi Rajkumar, Prasanna Kumar Kumaresan, Elizabeth Sherly, and Ruba Priyadharshini. 2023. Vel@ dravidianlangtech: Sentiment analysis of tamil and tulu. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, pages 211–216.

Bharathi B Ponsubash Raj R, Paruvatha Priya B. 2025. Justatalentedteam@dravidianlangtech 2025: A study of ml and dl approaches for sentiment analysis in code-mixed tamil and tulu texts. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Ratnavel Rajalakshmi, Srivarshan Selvaraj, Faerie Mattins, Pavitra Vasudevan, and Anand Kumar. 2023. HOTTEST: Hate and offensive content identification in tamil using transformers and enhanced STemming. *Comput. Speech Lang.*, 78(101464):101464.

Lavanya Sambath Kumar, Asha Hegde, Bharathi Raja Chakravarthi, Hosahalli 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*, pages 62–70, St. Julian's, Malta. Association for Computational Linguistics.

Koyyalagunta Krishna Sampath and M Supriya. 2024. Transformer based sentiment analysis on code mixed data. *Procedia Comput. Sci.*, 233:682–691.

Kogilavani Shanmugavadivel, Malliga Subramanian, Sanjai R, Mohammed Sameer, and Motheeswaran K. 2025. Beyond_tech@dravidianlangtech 2025: Political multiclass sentiment analysis using machine learning and neural network. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

E Sivasankar, Kalyanasundaram Krishnakumari, and P Balasubramanian. 2021. An enhanced sentiment dictionary for domain adaptation with multi-domain dataset in tamil language (esd-da). *Soft Computing*, 25:3697–3711.

K Sreeja and B Bharathi. 2025. Multimodal hate speech detection in dravidian languages. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Ippatapu Venkata Srichandra, Harish Vijay V, Pathange Omkareshwara Rao, and Premjith B. 2025. Deep learning approach for 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.

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