Sarcasm Detection in Dravidian Languages Using Bi-directional LSTM

Sulaksha B K[1,*], Shruthi Priyaa G K[2], Amritha P[3] and C. Jerin Mahibha[4]

,2,3,4[1] *Meenakshi Sundararajan Engineering College, Chennai, Tamil Nadu, India*

Abstract

Sarcasm is a form of speech in which one expresses a thing but intends the opposite, typically to express disdain, derision, or ridicule. Detecting sarcasm using machine learning is intricate due to its dependence on contextual nuances, lack of tone, and adverse subtlety, making it hard to interpret from text alone. This paper addresses this issue using a supervised learning model called bidirectional-Long Short Term Memory (LSTM) on Tamil and Malayalam text data to classify sarcastic texts, and saves the test results in a Comma Seperated Values (CSV) file. It uses tokenization and sequence padding for text preprocessing and is trained on a labeled dataset to capture semantic characteristics and features of the given text. For Detecting Sarcasm in ”Sarcasm Identification in Dravidian Languages in in Bi-Directional LSTM” the proposed best-performing LSTM model, achieved the Third position in the Tamil-English subtask (Macro-F1 Score: 0.72) and also in the Malayalam-English subtask (Macro-F1 Score: 0.74).

Keywords

Bi-directional LSTM, Machine Learning, Dravidian Languages, Sequence Padding, Natural Language Processing, Sarcasm Detection

1. Introduction

Sarcasm is a word derived from the Greek verb ”Sark'azein,” which means to speak bitterly. These words are often used in a humorous way to mock people [1]. Sarcasm requires some shared knowledge between speaker and audience; it is a profoundly contextual phenomenon. Most computational approaches to sarcasm detection, however, treat it as a purely linguistic matter, using information such as lexical cues and their corresponding sentiment as predictive features [2]. Detecting the sarcastic comments on social media has received much attention in the recent days, social media comments frequently use include positive words that represent negative attributes or characteristics.

Generally, if something is said like, "Wow, thanks a lot for arriving on time." it's easy to tell it's not sarcastic, this normal sentence sincerely expresses gratitude. But if the sentence is observed, the positive words "thanks a lot" expresses that individual is often late, it's a sign that the person is trying to be sarcastic. For Detecting Sarcasm in **"Sarcasm Identification in Dravidian Languages using Bi-Directional LSTM"** the proposed best-performing LSTM model, achieved the **Third position** in the Tamil-English subtask **(Macro-F1 Score: 0.72)** and also in the Malayalam-English subtask **(Macro-F1 Score: 0.74).** This paper talks about the project and research in various sections and subsections such as Related Works which talks about existing systems and previously present models, Dataset Description which provides a detailed description about the input, development and test datasets for both Malayalam and Tamil, Proposed Methodology which talks about the method and models used in this project, the Results section features the prediction dataset generated by the developed model and classification reports for both the Dravidian Languages, the Future Enhancement section addresses the ares of improvement in the project, and the project finally concludes with addressing the challenges present in the field and the References for this paper are cited and addressed.[3] [4]

Forum for Information Retrieval Evaluation, December 12-15, 2024, India.
bksulaksha@gmail.com (S. B. K); sprb0406@gmail.com (S. P. G. K); amriamritha03@gmail.com (A. P); jerinmahibha@msec.edu.in (C. J. Mahibha)

2. Related Works

S. M. Sarsam, H. Al-Samarraie, A. I. Alzahrani, B. Wright [5] had proposed a model to detect sarcasm in twitter using machine learning algorthms like Support Vector Machine (SVM). The review results revealed that Support Vector Machine (SVM) was the most commonly used Adapted Machine Learning Algorithms (AMLA) for sarcasm detection in Twitter. In addition, combining Convolutional Neural Network (CNN) and SVM was found to offer a high prediction accuracy. Mondher Bouazizi et al. [6] had come up with a pattern-based approach to detect sarcasm on Twitter. They had divided words into two classes: a first one referred to as *"CI"* containing words of which the content is important and a second one referred to as *"GFI"* containing the words of which the grammatical function is more important. If a word belongs to the first category, it is lemmatized; otherwise, it is replaced by certain expression. The approach reached an accuracy of 83.1% with a precision equal to 91.1%. Meriem et al. [7] had come up with a fuzzy approach to solve the task. This approach focused mainly on predicting the right label based on a measure known as the Sarcasm Score Measure, which calculates the measure of sarcasm based on which the prediction was made. This model had been implemented on two datasets: one is SemEval2014, and the other is the Bamman et al. dataset. They had attained an F1-score of 75.9 and 74.8 percent, respectively. Both binary and multi-class classifications were used in this task. The work of Santiago Castro et al. [8] explored the role of multimodality and conversational context in sarcasm detection and introduced a new resource to further enable research in this area. More specifically, the paper maked the following contributions: 1. Curate a new dataset, MUStARD, for multimodal sarcasm research with high-quality annotations, including both mutlimodal and conversational context features; 2. Exemplify various scenarios where incongruity in sarcasm is evident across different modalities, thus stressing the role of multimodal approaches to solve this problem; 3. To introduce several baselines and show that multimodal models are significantly more effective when compared to their unimodal variants; and 4. They also provided preceding turns in the dialogue which act as context information. The spawn of internet and communication technologies, distinctly the online social networks have modernized how people interact and communicate with each other digitally. People tend to express more on Social Media and as anonomous reveiws compared to real life interactions and communications Amir et al. [9] proposed the use of IndicBERT for the process of detecting sarcasm from social media text, the model effectively captured sarcasm cues and contextual information within the text S. Amir, B. C. Wallace, H. Lyu, P. C. M. J. Silv [10] . A pragmatic and intelligent model for sarcasm detection in social media text has been proposed by Mayank Shrivastava and Shishir Kumar. [11] which was based on Google BERT (Bidirectional Encoder Representations from Transformers) that can handle volume, velocity and veracity of data. Madhumitha M et al. had used different transformer models for detecting sarcasm from Tamil text M. M, K. Akshatra M, T. J, C. Mahibha, D. Thenmozhi, [12]. Agrawal et al. [13] had formulated the task of sarcasm detection as a sequence classification problem by leveraging the natural shifts in various emotions over the course of a piece of text. The proposed model by D. Jain, A. Kumar, G. Garg [14] is a hybrid of bidirectional long short-term memory with a softmax attention layer and Convolution Neural Network for real-time sarcasm detection. The methodology had used ELMo embedding based Convolutional Neural Network model, TF-IDF based Gaussian Naive Bayes classifier. D. Krishnan, J. M. C, T. Durairaj, [15] used the dataset provided by SemEval-2022 to discern and classify different types of irony within textual content. Kalaivani and Thenmozhi [16] had done sentiment analysis on the Dravidian-CodeMix-FIRE2021 dataset [17], where comments in 3 languages were trained: Tamil, Malayalam, and Kannada. They had used the pre-defined BERT model to perform this task.

3. Dataset Description

This research integrates datasets of Tamil and Malayalam texts, meticulously designed to advance sarcasm analysis models. These datasets encompass a broad spectrum of text types, including film reviews, social media comments, and general online interactions, and are annotated to differentiate

Table 1
Tamil Training and Development Dataset Sample

Text	Label
Apadilam nadakathu nadakavum koodathu.. Mass dialogue	Non-sarcastic
Avara Contol Pannunga Pls... Vera level expression... Thala...	Sarcastic
2:01 (sound missing) ennada ivlo kodi potu padam eduthurukeenga.	Non-sarcastic
Thala Ajith sir....mass konjam kammiya irukuravanga like pannunga	Non-sarcastic
Sun pictures nalla panringa THALAIVAR DHARISANAM	Sarcastic
ASURAN trailer .... Kollla massu.... asuran dhanush	Sarcastic

Table 2
Malayalam Training and Development Dataset Sample

Text	Label
Screenshot edukkan vannth njan	Sarcastic
Mollyhood is getting bigger and bigger	Non-sarcastic
Adukala oru aan kutiye palathum padipikum.....parasyam vannavar	Non-sarcastic
Sha rukh Khan nte fan padam pole undallo	Sarcastic
Waiting from die hard rajuvetan fan..	Non-sarcastic
Raju ettan fansinte watsapp group undenkil pls add me..	Non-sarcastic

between sarcastic and non-sarcastic texts. Texts were systematically collected from Tamil and Malayalam film review platforms, social media networks, and fan forums, ensuring comprehensive representation across various contexts where sarcasm, is conveyed. The model was trained with two datasets for each language :-

- Training Dataset
- Development Dataset

The datasets are categorized into two primary classes. Non-sarcastic texts communicate direct sentiments or opinions without employing irony. For instance, these may include well-wishes for an actor's opportunity in a film or positive feedback on a trailer's quality. In contrast, sarcastic texts utilize irony to convey sentiments that often contradict their literal meaning. Such texts may mock the exaggerated aspects of a film's presentation or deride its quality despite seemingly favorable remarks.

The datasets reveal that sarcastic remarks frequently use hyperbolic expressions and present stark contrasts with literal statements, creating complex challenges for the analysis. Table 1 and Table 2 are some of the example instances from the training and development dataset given for prediction of labels. The Dataset provided for this task is referenced from multiple datasets [18] [19] [20] [21] [22]

These annotated datasets are crucial for developing advanced machine learning models capable of effectively identifying and interpreting both sarcasm and nuanced sentiments in Tamil and Malayalam texts. The insights and tools derived from this research have substantial implications for automated sarcasm analysis applications, including social media monitoring, customer feedback evaluation, and film reviews.

By addressing the intricacies of sarcasm in both languages, this study makes a significant contribution to the broader field of natural language processing.

The dataset is categorized into two primary labels.

- Sarcastic
- Non-Sarcastic

Table 3
Tamil Test Dataset Sample

ID	Text
Id 07	Thalaivaa vaa vaa .....Thaa evanum kitta vara kudadhu
Id 08	Haha Sema documentary thetre ku varudhu polayae
Id 09	Surya Ku pair illa ya ?
Id 10	Kaithi trailer yen trending la varala
Id 11	Neenga ept padam edukka edukka than kathal athikam aga poguthu da
Id 12	Let's make it 30M , kabali, Kaala and now Petta

Table 4
Malayalam Test Dataset Sample

ID	Text
Id 30	Mammookka fans inu like adikkan ulla comment
Id 31	Now 5.2k dislikes Trailer varunnathine "Andha bayam irukkanam"
Id 32	Lucifer le item dance video song erakkan patvo... Illa le
Id 33	Pls support me pls My channel subscribe pls Pls Pls
Id 34	Prithviraj or Mammootty cheyyanda role aanu yenna
Id 35	Ippo penpidi kazhinju.! Kunju pidipikaan thudungayo? manasilayo yeda mone

4. Proposed Methodology

In this section, the methodology used for the complex task of detecting sarcasm in Dravidian languages(Tamil and Malayalam) is explored. The aim is to dissect the detailed process, emphasizing its different phases, and clarify how each step plays a significant role in achieving the main objective. The proposed strategy utilizes Natural Language Processing (NLP) techniques and machine learning models to address this linguistic challenge.

4.1. Data Preparation

The Dataset Preparation involves loading and combining the training and development dataset and extracting the text and its labels which is used for prediction.

4.1.1. Loading and Combining the Dataset

The training, development and testing datasets are imported and loaded in the program. The training dataset is combined with the development dataset to create a larger dataset, labels merged which helps improve the model's ability to generalize.

4.1.2. Text and Label Extraction

The texts in the dataset is tokenized using the tokenizer class in the keras model which converts the text to integer sequences. It is used to create a dictionary of the most frequent words which is padded to a uniform length which is necessary for input into neural networks. The extracted labels are then encoded into numeric format using Label Encoder. This step converts labels into binary format for classification.

4.2. Model Design

The model begins with the embedding layer which converts input tokens into dense vectors. It captures their semantic meanings. This is followed by three Bi-Directional Layers, each processing in both

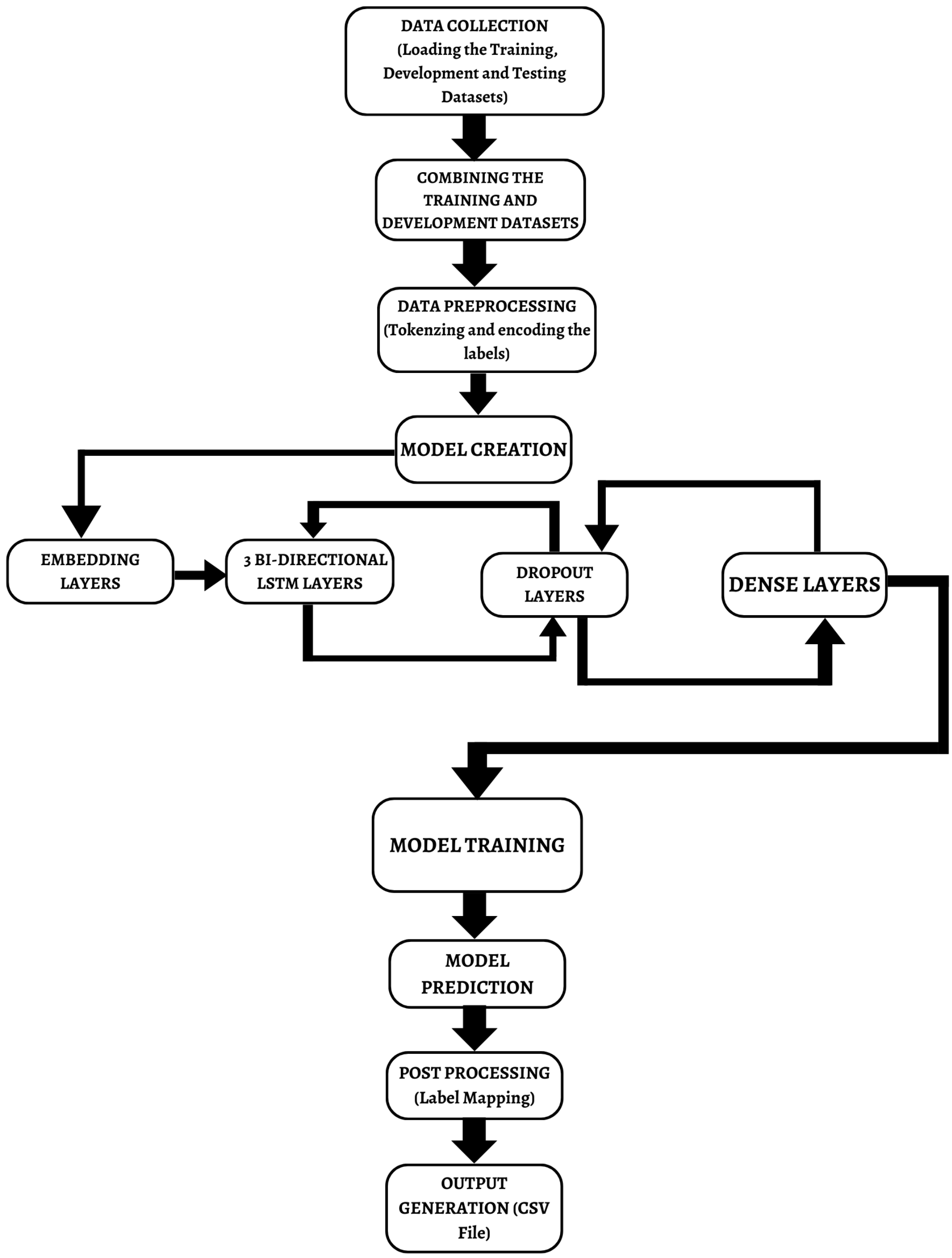

Figure 1: Model Architecture

forward and backward directions. The final Bi-Directional LSTM layer integrates the representations before passing the output to the dense layer.

4.2.1. Embedding Layer

It helps the deep learning models to understand real world data domains very effectively. It maps each word index to 256-dimensional vector which helps the model understand the semantic relationships between the words. These high-level embeddings facilitate more precise and meaningful analysis, thus improving the model's overall performance.

4.2.2. Bi-Directional LSTM Layers

Bidirectional LSTM or Bi-LSTM is used for a sequence model. It contains two LSTM layers, for processing input in both forward and backward directions.The bidirectional LSTM is better when compared with unidirectional LSTM.Three Bi-Directional LSTM Layers are used for capturing forward and backward contextual information from text. The first layer consists of 128 units, second layer consists of 64 units and the third layer consists of 32 units. The first two layers returns sequences and the third layer returns the final output.

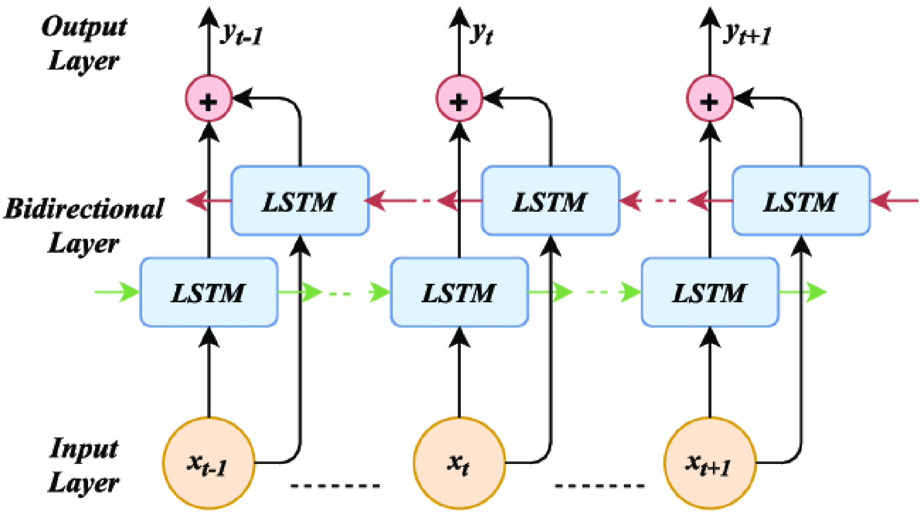

Figure 2: Architecture of Bi-Directional LSTM

4.2.3. Dense Layer

The dense layer with 32 units and ReLU activation is added to reduce the dimensionality of the output after the LSTM layers. The final dense layer with 1 unit and sigmoid activation is added for binary classification. The combination of these layers helps the model generalize by balancing feature extraction and classification.

4.2.4. Dropout Layer

A dropout layer is added after each LSTM and Dense layers to prevent overfitting by reducing 50% of the neurons during training. This helps the model become too relient on any specific set of neurons. By applying Dropout, the model becomes more generalized, thereby improving the performance on unseen data.

4.3. Model Compilation and Training

The model is compiled using **Loss function:Binary Crossentropy, Optimizer: optimizer and Metric:Accuracy.**The model is trained using combined dataset of padded-sequences-combined for input and label-encoded-combined for target labels. 10% of the training data is reserved for validation.

Training is performed over 10 epochs and 32 units(batch size) which controls the number of samples to be processed before updating the model's internal parameters.

4.4. Predictions on Test Data

The labels in the test data are predicted using the trained model using the padded-sequences-test. The predictions are threshold at 0.5 to convert them into binary classes(0 or 1). The predicted labels are mapped to their original labels(Sarcastic and Non-Sarcastic). The predicted labels are saved to the CSV file and classification report is generated. Together, tokenization, model training, and meticulous evaluation helped to unravel the intricacies of sarcasm detection in Dravidian languages.

Table 5
Tamil Prediction Dataset Sample

ID	Predicted Label
Id 07	Non-Sarcastic
Id 08	Non-sarcastic
Id 09	Sarcastic
Id 10	Sarcastic
Id 11	Sarcastic
Id 12	Non-sarcastic

Table 6
Malayalam Prediction Dataset Sample

ID	Predicted Label
Id 30	Sarcastic
Id 31	Non-sarcastic
Id 32	Non-sarcastic
Id 33	Sarcastic
Id 34	Non-sarcastic
Id 35	Non-sarcastic

5. Results

In Tables 5 and 6. a sample of the predicted dataset by the model is given. The dataset contains two columns that is ID and Predicted Labels. In Tables 7 and 8, the performance for both models across both languages are presented. From the classification reports, it is observed that the key statistics related to the performance of the model is accuracy. It could be observed that the accuracy of 0.78 and 0.82 is achieved in the Predicted Tamil and Malayalam dataset models respectively.

The classification result plays a major role in identifying the strengths and weaknesses of the model which helps in fine-tuning process and can thus improve the performance of the model. The performance of the proposed models is examined using a range of evaluation metrics, with a primary focus on F1-Score, accuracy, recall, macro-averaged F1-score, and weighted average F1-score. The organizers thoughtfully provided test data for both Dravidian languages, which served as the foundation for the model evaluation.

The results of the Sarcasm detection task from the organizers is shown in Table 5. The model has achieved excellence by securing the third rank in both Tamil and Malayalam classification procedures. Using the bi-directional LSTM model, the labels were predicted for the comments given in the dataset. It provided an accuracy of 0.78 and 0.82 for the Tamil and Malayalam datasets, respectively. Macro-F1 scores of 0.72 and 0.45 were achieved in the Tamil and Malayalam datasets, respectively. It is evident

Table 7
Classification Report (Tamil)

	Precision	Recall	F1-Score	Support
Non-Sarcastic	0.85	0.84	0.84	4621
Sarcastic	0.58	0.62	0.60	1717
Accuracy			0.78	6338
Micro Avg	0.72	0.73	0.72	6338
Weighted Avg	0.78	0.78	0.78	6338

Table 8
Classification Report (Malayalam)

	Precision	Recall	F1-Score	Support
Non-Sarcastic	0.82	1.00	0.90	2314
Sarcastic	0.50	0.00	0.01	512
Accuracy			0.82	2826
Micro Avg	0.66	0.50	0.45	2826
Weighted Avg	0.76	0.82	0.74	2826

Table 9
Results

Tasks	Model	Runs	Macro-Fi	Rank
Tamil Dataset	Bi-Directional LSTM	2	0.72	3
Malayalam Dataset	Bi-Directional LSTM	2	0.74	3

that out of 6339 comments, 1717 comments are sarcastic and 4621 comments are non-sarcastic in the Tamil dataset. Similarly, in the Malayalam dataset, out of 2827 comments, 512 are Sarcastic and 2314 are Non-sarcastic.

6. Future Enhancements

The computer science field is not static, it is subjected to be dynamic. The technology which is popular today becomes outdated the next day itself. The future enchancement refer to the improvements which can be done in the future stages of the project to increase the accuracy, effciency and performance metrics.

In this sarcasm detection model, the enhancement could be done by using the advanced method SMOTE (Synthetic Minority Over-Sampling Technique) or by oversampling the minority class (sarcastic comments) and undersampling the majority class (non-sarcastic comments) especially in the malayalam dataset. Pre-trained language models like BERT (Bidirectional Encoder) or mBERT (Multilingual BERT) could also be incorporated which can enhance the model's ability to detect the sarcastic comments.

Another enhancement could be adding features to the model such as user behavior, replies in conversation can be useful for detecting the sarcastic comments, which are contextual features. Pre-trained sarcasm detection models on English dataset could also be fine-tuned and used for Tamil and Malayalam languages.

7. Conclusion

Detecting sarcasm in Dravidian languages (Tamil and Malayalam), presents challenges due to the languages' linguistic diversity, cultural differences, and complex sentence structures. Using advanced deep learning models like Bidirectional LSTMs, along with specialized tokenization and embedding techniques, has led to significant progress in sarcasm detection. Sarcasm in Tamil and Malayalam often hinges on deeper contextual cues, including tone, cultural context, and socio-political factors. Incorporating language-specific factors, grammers etc., and improving the model's ability to understand context can significantly enhance accuracy of the model. Moreover, extending the work to include other languages promises to broaden the scope and applicability of the methodology . In conclusion, the research represents a valuable step forward in the field of sarcasm detection for Dravidian languages(Tamil and Malayalam). By comparing the proposed approach and findings with prior studies, it could contribute to the ongoing discourse and innovation in this area, helping to drive the development of more precise and robust sarcasm detection systems.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] A. Reyes, P. Rosso, T. Veale, A multidimensional approach for detecting irony in twitter, Language resources and evaluation 47 (2013) 239–268.

[2] D. Bamman, N. Smith, Contextualized sarcasm detection on twitter, in: proceedings of the international AAAI conference on web and social media, volume 9, 2015, pp. 574–577.

[3] V. Indirakanth, D. Udayakumar, T. Durairaj, B. Bharathi, Sarcasm identification of dravidian languages (malayalam and tamil)., in: FIRE (Working Notes), 2023, pp. 327–335.

[4] B. R. Chakravarthi, S. N, B. B, N. K, T. Durairaj, R. Ponnusamy, P. K. Kumaresan, K. K. Ponnusamy, C. Rajkumar, Overview of sarcasm identification of dravidian languages in dravidiancodemix@fire-2024, in: Forum of Information Retrieval and Evaluation FIRE - 2024, DAIICT , Gandhinagar, 2024.

[5] S. M. Sarsam, H. Al-Samarraie, A. I. Alzahrani, B. Wright, Sarcasm detection using machine learning algorithms in twitter: A systematic review, International Journal of Market Research 62 (2020) 578–598.

[6] M. Bouazizi, T. O. Ohtsuki, A pattern-based approach for sarcasm detection on twitter, IEEE Access 4 (2016) 5477–5488.

[7] A. B. Meriem, L. Hlaoua, L. B. Romdhane, A fuzzy approach for sarcasm detection in social networks, Procedia Computer Science 192 (2021) 602–611.

[8] S. Castro, D. Hazarika, V. Pérez-Rosas, R. Zimmermann, R. Mihalcea, S. Poria, Towards multimodal sarcasm detection (an _obviously_ perfect paper), arXiv preprint arXiv:1906.01815 (2019).

[9] B. Sulaksha, S. Krishnaveni, I. Steeve, et al., Sis@ lt-edi-2023: Detecting signs of depression from social media text, in: Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion, 2023, pp. 131–137.

[10] S. Amir, B. C. Wallace, H. Lyu, P. C. M. J. Silva, Modelling context with user embeddings for sarcasm detection in social media, arXiv preprint arXiv:1607.00976 (2016).

[11] M. Shrivastava, S. Kumar, A pragmatic and intelligent model for sarcasm detection in social media text, Technology in Society 64 (2021) 101489.

[12] M. M, K. Akshatra M, T. J, C. Mahibha, D. Thenmozhi, Sarcasm detection in dravidian languages using transformer models (2023).

[13] A. Agrawal, A. An, M. Papagelis, Leveraging transitions of emotions for sarcasm detection, in:

Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, pp. 1505–1508.

[14] D. Jain, A. Kumar, G. Garg, Sarcasm detection in mash-up language using soft-attention based bi-directional lstm and feature-rich cnn, Applied Soft Computing 91 (2020) 106198.

[15] D. Krishnan, J. M. C, T. Durairaj, GetSmartMSEC at SemEval-2022 task 6: Sarcasm detection using contextual word embedding with Gaussian model for irony type identification, in: G. Emerson, N. Schluter, G. Stanovsky, R. Kumar, A. Palmer, N. Schneider, S. Singh, S. Ratan (Eds.), Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022), Association for Computational Linguistics, Seattle, United States, 2022, pp. 827–833. URL: https://aclanthology.org/2022.semeval-1.114. doi:10.18653/v1/2022.semeval-1.114.

[16] A. Kalaivani, D. Thenmozhi, Multilingual sentiment analysis in tamil, malayalam, and kannada code-mixed social media posts using mbert, FIRE (Working Notes) (2021) 1020–1028.

[17] C. J. Mahibha, S. Kayalvizhi, D. Thenmozhi, Sentiment analysis using cross lingual word embedding model (2021).

[18] N. Sripriya, T. Durairaj, K. Nandhini, B. Bharathi, K. K. Ponnusamy, C. Rajkumar, P. K. Kumaresan, R. Ponnusamy, C. Subalalitha, B. R. Chakravarthi, Findings of shared task on sarcasm identification in code-mixed dravidian languages, FIRE 2023 16 (2023) 22.

[19] B. R. Chakravarthi, N. Sripriya, B. Bharathi, K. Nandhini, S. C. Navaneethakrishnan, T. Durairaj, R. Ponnusamy, P. K. Kumaresan, K. K. Ponnusamy, C. Rajkumar, Overview of the shared task on sarcasm identification of dravidian languages (malayalam and tamil) in dravidiancodemix, in: Forum of Information Retrieval and Evaluation FIRE-2023, 2023.

[20] B. R. Chakravarthi, Hope speech detection in youtube comments, Social Network Analysis and Mining 12 (2022) 75.

[21] B. R. Chakravarthi, A. Hande, R. Ponnusamy, P. K. Kumaresan, R. Priyadharshini, How can we detect homophobia and transphobia? experiments in a multilingual code-mixed setting for social media governance, International Journal of Information Management Data Insights 2 (2022) 100119.

[22] B. R. Chakravarthi, N. Sripriya, B. Bharathi, K. Nandhini, S. Chinnaudayar Navaneethakrishnan, T. Durairaj, R. Ponnusamy, P. K. Kumaresan, K. K. Ponnusamy, C. Rajkumar, Overview of the shared task on sarcasm identification of Dravidian languages (Malayalam and Tamil) in DravidianCodeMix, in: Forum of Information Retrieval and Evaluation FIRE - 2023, 2023.