

Investigation of Machine Learning and Transformer Models for Sarcasm Detection in Dravidian Languages

Malliga Subramanian, Ananthakumar S, Deepiga P, Dharshini S, Kogilavani S V¹

¹Department of Computer Science and Engineering

²Kongu Engineering College Erode Tamil Nadu India

Abstract

Sarcasm detection in natural language is a critical task for sentiment analysis, particularly in resource-constrained languages like Malayalam, a Dravidian language. We secured 6th place in the Sarcasm Identification of Dravidian Languages (Malayalam) track at DravidianCodeMix@FIRE-2024. The complexity of sarcasm, which often relies on context, tone, and cultural understanding, poses unique challenges for machine learning models. This study investigates the effectiveness of various machine learning and deep learning techniques in identifying sarcasm in Malayalam text. We employ a diverse set of models, including RoBERTa, Convolutional Neural Networks (CNN), Multilayer Perceptron (MLP), Gated Recurrent Units (GRU), Recurrent Neural Networks (RNN), Random Forests (RF), Hidden Markov Models (HMM), Logistic Regression (LR), K-Nearest Neighbors (KNN), and Gaussian Mixture Models (GMM).

1. Introduction

Sarcasm, characterized by the use of irony to mock or convey contempt, is one of the most complex linguistic phenomena to detect in natural language processing (NLP) [2]. In human communication, sarcasm often relies on nuanced cues such as tone, facial expressions, or context that are difficult to interpret in written text. Its proper identification is crucial for sentiment analysis, opinion mining, and emotion recognition [9]. The challenge of sarcasm detection becomes even more pronounced in less-resourced languages, where annotated datasets and linguistic resources are limited. [4] Malayalam, a Dravidian language spoken primarily in the Indian state of Kerala, presents unique difficulties for sarcasm detection due to its rich morphology, complex grammatical structure, and diverse expressions. [1] In recent years, several machine learning and deep learning techniques have been employed to address sarcasm detection [6], but their application in Malayalam is relatively underexplored. Previous studies have predominantly focused on widely spoken languages like English, leaving a significant gap in NLP research for regional languages like Malayalam. The limited availability of annotated sarcasm datasets in Malayalam further exacerbates this problem, making it difficult to train robust models for sarcasm identification. In this study, we focus on sarcasm identification in Malayalam using a variety of machine learning and deep learning models. [1] provides a comprehensive overview of sarcasm detection techniques across various languages, including applications for Dravidian languages. We explore transformer-based model like RoBERTa, which have demonstrated state-of-the-art performance in NLP tasks by leveraging contextual embeddings. Additionally, we apply deep learning architectures such as Convolutional Neural Networks (CNN), Multilayer Perceptron (MLP), Gated Recurrent Units (GRU), and Recurrent Neural Networks (RNN), which excel in capturing semantic and sequential information from text. To benchmark the performance of traditional machine learning approaches, we also employ models like Random Forests (RF), Hidden Markov Models (HMM), Logistic Regression (LR), K-Nearest Neighbors (KNN), and Gaussian Mixture Models (GMM). This study aims to evaluate the effectiveness of these models in detecting sarcasm in Malayalam, identifying key challenges, and comparing the performance of different approaches. By exploring a diverse range of models, we seek to provide insights into the strengths and limitations of machine learning and deep learning techniques for

Forum for Information Retrieval Evaluation, December 12-15, 2024, India

✉ deepigap.22cse@kongu.edu (M. S. A. S. D. P. D. S. K. S. V)

🌐 <https://github.com/Deepikagowtham/Sarcasm-Detection.git> (M. S. A. S. D. P. D. S. K. S. V)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

sarcasm detection in low-resource languages like Malayalam, ultimately contributing to the development of more robust NLP tools for Dravidian languages.

2. Literature Survey

Sarcasm detection is a crucial area of research in natural language processing (NLP), especially for sentiment analysis, opinion mining, and emotion recognition. The survey [1] provides a comprehensive overview of sarcasm detection techniques across various languages, including applications for Dravidian languages. The survey [2] explores sarcasm detection using machine learning techniques applicable to microblogs and social media data. The survey [3] findings of the shared task on offensive language identification in Dravidian languages. Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages. This shared task includes sarcasm detection in Malayalam. While substantial progress has been made in sarcasm identification for English and other well-resourced languages, the research for low-resource languages, including Dravidian languages such as Malayalam, remains limited. The survey [4] Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP). The survey [5] gives proceedings of the Second Workshop on Multilinguality and Code Switching (MultiLingCode). Discusses sarcasm detection in Tamil, which shares linguistic similarities with Malayalam. The survey [6] building a dictionary of affective words for a sentiment analysis of online messages. Journal of Language Resources and Evaluation, 50(3), 665-691. This paper focuses on sentiment analysis, an underlying concept for sarcasm detection in Malayalam. This paper discusses the role of personalized contexts in detecting sarcasm, useful for Dravidian languages. the survey [7] gives proceedings of the Third Workshop on Computational Approaches to Linguistic Code-Switching. This work explores the classification of code-mixed Dravidian languages, applicable to sarcasm detection in Malayalam. The complexity of sarcasm detection arises from its reliance on context, tone, and cultural nuances, which make it a non-trivial task for machine learning and deep learning models. The survey [8] Proceedings of the International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things. Discusses the role of sarcasm in sentiment analysis, relevant to detecting sarcasm in Malayalam. The survey [9] presents machine learning techniques for sarcasm detection, useful for Dravidian languages. The survey [10] gives proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages. Offers insights into sentiment and sarcasm detection in code-mixed Indian languages, including Malayalam. The survey [11] show their findings highlight the effectiveness of different modeling approaches in understanding sarcasm in textual data. The survey [12] provides an overview of the models submitted by participants for the task of sarcasm identification in Dravidian languages, as presented in DravidianCodeMix@FIRE-2024. It highlights the methodologies employed, the diversity of approaches, and the overall contributions of each submission, aiming to enhance understanding and improve future efforts in this area of research.

2.1. Sarcasm Detection in English and Other Major Languages

The earliest approaches to sarcasm detection relied on rule-based and lexical analysis techniques, which were primarily centered on identifying specific syntactic or semantic patterns. Davidov et al. (2010) introduced a semi-supervised sarcasm detection technique that leveraged features like patterns, punctuation, and n-grams from Twitter data in English [10]. These methods, though effective in specific scenarios, struggled with generalization due to the nuanced nature of sarcasm, which is highly context-dependent. As machine learning models advanced, supervised learning methods such as Support Vector Machine (SVM), Random Forests (RF), and Logistic Regression (LR) became more widely used in sarcasm detection tasks [8]. These models often relied on hand-crafted features, including n-grams, sentiment lexicons, and part-of-speech tags. While these methods provided improvements, their reliance on manually engineered features limited their ability to capture deeper contextual and linguistic nuances essential for sarcasm detection. More recently, transformers like BERT (Bidirectional Encoder Representations from Transformers) and its variants (e.g., RoBERTa, ALBERT) have set new benchmarks

in NLP tasks by utilizing self-attention mechanisms to understand complex word relationships in context. These models have achieved state-of-the-art performance in sarcasm detection for English due to their ability to leverage contextual embeddings, which capture intricate meanings within sentences.

2.2. Sarcasm Detection in Dravidian Languages

Despite the advances in English sarcasm detection, research in sarcasm identification for Dravidian languages like Malayalam has been sparse. Dravidian languages are morphologically rich and exhibit complex syntactic structures, making sarcasm detection particularly challenging [5]. The lack of large annotated datasets for these languages adds to the difficulty of building reliable models. Most research in Malayalam has focused on sentiment analysis and emotion detection, often using shallow machine learning models like Naive Bayes, SVM, and RF. For instance, Kumar et al. (2020) explored sentiment analysis in Tamil and Malayalam using traditional machine learning techniques [3], demonstrating that these languages require more advanced methods for sarcasm detection. The reliance on hand-crafted features limits the effectiveness of these models in capturing the full complexity of sarcasm.

2.3. Deep Learning and Transformer Models for Malayalam

Given the success of deep learning models in other languages, there is growing interest in applying these methods to sarcasm detection for Malayalam. Models like CNN, Multilayer Perceptron (MLP), and Gated Recurrent Units (GRU) have shown promise in handling the intricacies of text classification tasks by capturing semantic and sequential information in the data. [11] However, the application of these models to Malayalam sarcasm detection is still underexplored due to the limited availability of labeled data. Transformer-based models such as RoBERTa, and GPT have revolutionized NLP tasks by capturing complex relationships in the text through self-attention mechanisms [7]. These models could significantly improve sarcasm detection in Malayalam by leveraging transfer learning, where models pre-trained on large datasets (such as multilingual BERT) are fine-tuned for Malayalam-specific tasks. Preliminary research in sentiment analysis for Dravidian languages using transformer models suggests that they have great potential for sarcasm detection as well, provided there are enough annotated resources for fine-tuning.

2.4. Challenges and Future Directions

One of the primary challenges in sarcasm detection for Malayalam is the lack of large-scale annotated datasets. Manual annotation of sarcasm is resource-intensive due to the nuanced nature of the task. Additionally, Malayalam exhibits rich morphology, including inflectional changes that make text normalization and feature extraction challenging. Future research should focus on building more extensive annotated corpora and developing hybrid approaches that combine traditional machine learning models with deep learning architectures to better capture the complexity of sarcasm. In conclusion, while sarcasm detection for Malayalam and other Dravidian languages is in its early stages, there is significant potential to improve the task by adopting modern deep learning techniques and transfer learning approaches.

3. Materials and Methods

3.1. Dataset Description

The dataset for sarcasm identification in Malayalam consists of social media posts and online comments, annotated to classify each text as either sarcastic or non-sarcastic. The texts are sourced from platforms like Twitter and Facebook, where sarcasm is commonly used. Each post is labeled based on its contextual meaning and linguistic cues. Sarcastic texts (labeled as 1) convey an ironic or contradictory meaning compared to their literal interpretation, while non-sarcastic texts (labeled as 0) reflect straightforward expressions. The dataset is preprocessed to remove special characters, URLs, and stop-words, and it

	Text	labels
0	Screenshot edukkan vannth njan മാത്രമാണോ	Sarcastic
1	നമ്മുടെ അനു സിത്താര ചേച്ചി ഇങ്ങനെ വരുന്നത് നോക്ക...	Sarcastic
2	Mollyhood is getting bigger and bigger	Non-sarcastic
3	Ho aaa BGM. Mammookka ithu oru pwoli pwolikkum	Non-sarcastic
4	Enthaale, sambhavam puraanam aanelum backgroun...	Sarcastic

Figure 1: sarcastic or Non-Sarcastic

may include features like punctuation and polarity to help models better understand the underlying sarcasm. Models trained on this dataset are evaluated on metrics such as accuracy, precision, recall, and F1-score to assess their performance in distinguishing between sarcastic and non-sarcastic Malayalam texts.

3.2. Pre-processing

Pre-processing for sarcasm identification in Malayalam involves a series of steps designed to clean and standardize the text, preparing it for meaningful analysis. Given the language's rich morphology and frequent code-switching with English, the first step is data cleaning, which involves removing special characters, URLs, and irrelevant stop-words that do not contribute to the sarcastic content. This is particularly important when handling mixed-language text, where Malayalam and English are interspersed. Following data cleaning, text normalization is performed to standardize the words. Tokenization is the process of breaking down the text into smaller units, such as words or sub-words, which is crucial for understanding Malayalam's script and grammar. Stemming and lemmatization are applied to reduce words to their root forms, which helps in managing the language's complex inflectional morphology. For example, different forms of the same word can be converted to a common root, making it easier for models to understand the core meaning. Punctuation, which plays a significant role in sarcasm detection, is deliberately retained during pre-processing. Elements like exclamation marks, ellipses, or question marks can alter the tone of a statement and often indicate sarcasm, so they are treated as features rather than noise. Additionally, handling agglutination is crucial in Malayalam, as the language frequently combines multiple morphemes into a single word, creating compound words. By breaking these down into individual morphemes or identifying the root forms and affixes, the system can better interpret the word's meaning in context. These comprehensive pre-processing steps ensure the data is clean, normalized, and well-structured, laying the groundwork for effective sarcasm identification.

3.3. Feature Extraction

Feature extraction for sarcasm identification in Malayalam involves techniques aimed at capturing linguistic patterns and contextual nuances that distinguish sarcastic expressions. The process begins with n-gram features, such as unigrams, bigrams, and trigrams, which help identify word sequences that may indicate sarcasm when analyzed in context, for instance, phrases where the literal meaning contrasts with the intended tone. TF-IDF (Term Frequency-Inverse Document Frequency) is another important approach, assigning weights to words based on their frequency and uniqueness, highlighting terms that might carry ironic or sentimentally charged meanings. Word embeddings, like Word2Vec and BERT, enhance the understanding of sarcasm by providing dense vector representations of words, with Word2Vec capturing semantic similarity and BERT offering contextual embeddings that consider surrounding words. Sentiment and polarity features further contribute by detecting contrasts between the expressed sentiment and actual intent, where a positive word used in a negative scenario might signal sarcasm. Part-of-speech (POS) tagging adds a syntactic dimension, helping to identify sarcastic patterns through the use of specific adjectives or unexpected word combinations. Given Malayalam's

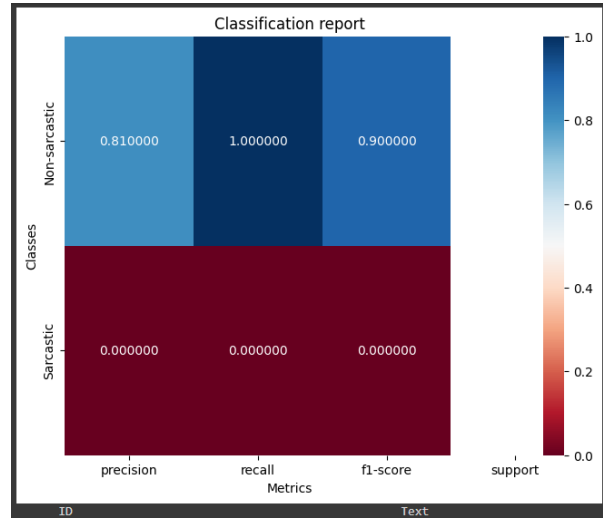


Figure 2: MLP

morphological richness, handling agglutination by decomposing complex words into their base forms and grammatical components is also essential. Additionally, recognizing idiomatic phrases and punctuation marks, such as exclamation points or ellipses, plays a significant role, as these elements often carry ironic undertones. Together, these techniques allow the model to capture both the linguistic features and contextual cues necessary for accurately detecting sarcasm in the text.

Table 1
Performance Table of our Models for Sarcasm Detection

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
KNN	56	64	63	64
SVM	77	83	80	83
DT	73	68	65	66

3.4. Proposed Classifiers

For sarcasm identification in Malayalam, performance metrics are used to evaluate how well the model distinguishes between sarcastic and non-sarcastic text. Accuracy measures the percentage of correctly classified examples but may not fully reflect performance when sarcasm is rare. Precision evaluates how many of the texts predicted as sarcastic are truly sarcastic, while Recall measures the model's ability to identify actual sarcastic texts, minimizing missed cases (false negatives). The F1-Score combines precision and recall into a balanced metric, useful when sarcasm is less frequent. A confusion matrix provides a detailed view of the model's errors, showing true positives, false positives, true negatives, and false negatives. These metrics together offer a comprehensive understanding of the model's strengths and weaknesses.

4. Dataset

The Training Dataset consists of 13,188 instances, including 2,499 categorized as sarcastic and 10,689 as non-sarcastic. In contrast, the Test Dataset includes 2,826 instances, with 521 labeled as sarcastic and 2,305 as non-sarcastic. Together, these datasets provide a robust foundation for training models to detect sarcasm in Malayalam text. The significant number of non-sarcastic examples in both datasets is crucial for effectively evaluating the model's performance during testing.

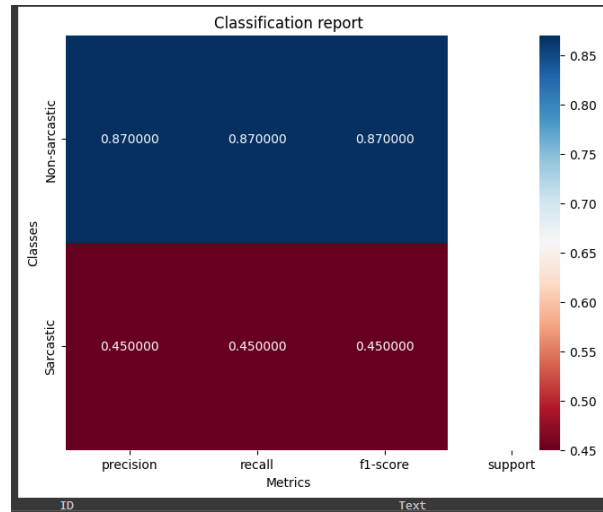


Figure 3: LSTM

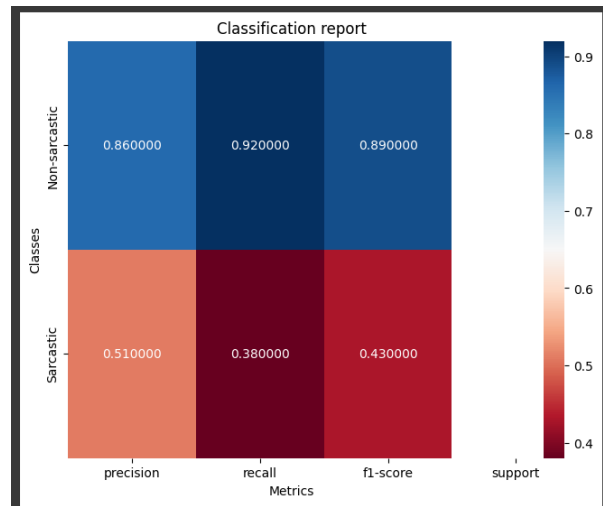


Figure 4: CNN

5. Conclusion and Future Work

Sarcasm identification in Malayalam, a Dravidian language, poses challenges due to its rich morphology, syntactic complexity, and frequent code-switching with English. Utilizing advanced classifiers like RoBERTa, CNNs, and GRUs alongside traditional models such as Random Forests and Logistic Regression, substantial progress has been made in detecting sarcasm, with models effectively capturing linguistic cues and contextual nuances. Pre-processing techniques like tokenization, sentiment analysis, and word embeddings have contributed to improved performance. However, further optimization is required to fully capture the complexity of sarcastic expressions in Malayalam. Future work could focus on fine-tuning advanced language models for Malayalam, leveraging cross-lingual models from other Dravidian languages, and building larger, diverse datasets from social media, including code-mixed content. Multimodal approaches incorporating text with audio or visual cues could further enhance sarcasm detection. Additionally, incorporating cultural and idiomatic knowledge into the models would provide a deeper understanding of sarcasm specific to Malayalam, enabling more accurate predictions.

Declaration on Generative AI

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

6. References

- [1] Joshi, A., Bhattacharyya, P., and Carman, M. J. (2017). Automatic Sarcasm Detection: A Survey. *ACM Computing Surveys (CSUR)*, 50(5), 1-22.
- [2] Mukherjee, S., and Bala, P. K. (2017). Sarcasm detection in microblogs using Naïve Bayes and fuzzy clustering. *Journal of Computer Science*, 13(5), 141-150.
- [3] Chakravarthi, B. R., Muralidaran, V., and Priyadharshini, R. (2021).
- [4] Ghosh, D., and Veale, T. (2017). Magnets for Sarcasm: Making Sarcasm Detection Timely, Contextual, and Very Personal.
- [5] Annamalai, R., and Muralidaran, K. (2021). Cross-lingual Sarcasm Detection for Code-Mixed Tamil-English Text.
- [6] Cruz, F. L., Troyano, J. A., Pontes, B., and Vallejo, C. G. (2016).
- [7] Chakravarthi, B., Priyadharshini, R., Kumar, M. A., and Jose, N. (2022). Dravidian Code-Mixed Text Classification using Transformer-based Models.
- [8] Kumar, V., and Garg, S. (2019). Sarcasm Detection and Its Impact on Sentiment Analysis: A Review.
- [9] Swain, M. Rajendran, A., and Singh, A. K. (2017). A comparison of machine learning techniques for sarcasm detection. *Journal of Computer Science and Technology*, 32(4), 729-744.
- [10] Chakravarthi, B. R., Kumar, S. A., Priyadharshini, R., and Muralidaran, V. (2020). Sentiment analysis of code-mixed Indian languages: An overview of SAIL Code-mixed shared task at ICON-20.
- [11] Kumar, A., Garg, G. (2019). Empirical study of shallow and deep learning models for sarcasm detection using context in benchmark datasets. *Journal of ambient intelligence and humanized computing*, 1-16.
- [12] Chakravarthi, B. R., N, S., B, B., K, N., Durairaj, T., Ponnusamy, R., Kumaresan, P. K., Ponnusamy, K. K., Rajkumar, C. (2024). Overview of sarcasm identification of Dravidian languages in DravidianCodeMix@FIRE-2024. In *Forum of Information Retrieval and Evaluation FIRE - 2024*. DAICT, Gandhinagar.