Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 2, No.3 : 2024
ISSN : **1906-9685**

SENTIMENT ANALYSIS IN LOW RESOURCE LANGUAGE: EXPLORING BERT, MBERT, XLM-R, AND RNN ARCHITECTURES TO UNDERPIN THE DEEP LANGUAGE UNDERSTANDING

Anitha R Research Scholar, Dept. of Futures Studies University of Kerala, Thiruvananthapuram
Email: anithathilak1717@gmail.com
K S Anil Kumar University of Kerala, Thiruvananthapuram Email: ksanilksitm@gmail.com

***Abstract*---**we investigate the sentiment analysis performance of multiple machine learning models (RNN, BERT, XLM-R, and mBERT) on a distinct dataset consisting of code-mixed, Malayalam, and Manglish (Malayalam written in Latin script). 22,449 entries make up the dataset, which was gathered through surveys and social media from 120 constituencies in Kerala. One of three attitude labels—positive, negative, or neutral—is appended to each text. To make sure the dataset was suitable for machine learning, a thorough preprocessing process that included tokenisation, normalisation, and lemmatisation was carried out. Using this dataset, we trained and assessed transformer-based architectures and RNN models. The results show that while RNN lags behind with an accuracy of 34%, BERT and XLM-R perform much better than typical RNN models, obtaining accuracies of 68% and 65%, respectively. These models were further improved by fine-tuning, with BERT achieving a 71% fine-tuned accuracy. The reliability of transformer-based models for managing intricate linguistic patterns in low-resource languages like Malayalam is confirmed by this study. With possible ramifications for social media surveillance, public opinion mining, and other areas, the findings provide insights into the application of sophisticated sentiment analysis algorithms for regional languages..

Keywords: Sentiment Analysis, RNN, BERT, XLM-R, mBERT

I. Introduction

Sentiment analysis in Malayalam text has gained significant attention due to the widespread use of social media and online platforms where users express their opinions and sentiments in their native language. The morphological richness and agglutinative nature of Malayalam, a Dravidian language basically spoken in the Indian state of Kerala and several neighboring regions, pose special obstacles for sentiment analysis. Sentiment analysis in Malayalam texts is essential for several applications, including public opinion mining, social media monitoring, and customer feedback analysis. Due to the increasing usage of social media and other online venues where people may express their ideas and sentiments in their local language, sentiment analysis in Malayalam texts has received a lot of attention. The morphological richness and agglutinative nature of Malayalam, a Dravidian language [24] predominantly spoken in the Indian state of Kerala and several neighboring regions, pose special obstacles for sentiment analysis. Sentiment analysis in Malayalam texts is essential for several applications, including public opinion mining, social media monitoring, and customer feedback analysis. Because Malayalam has so many different dialects and idioms, sentiment analysis needs models that can capture subtleties in language usage. The morphological differences and contextual ambiguities inherent in Malayalam pose a challenge to traditional sentiment analysis techniques, hence requiring the investigation of sophisticated natural language processing (NLP) models. Furthermore, sentiment analysis in Malayalam advances more general computational linguistics and natural language processing research, especially in the creation of language-specific models that are transferable to other Dravidian languages. In addition to improving our knowledge of sentiment dynamics in regional languages, this research makes it easier to automate the analysis of user-generated content on digital platforms with greater accuracy and cultural sensitivity.

In this study, we investigate how well a number of cutting-edge models including RNN, BERT(Bidirectional Encoder Representations from Transformers),XLM-R, and mBERT achieve these goals and improve Malayalam text sentiment analysis capabilities. We hope to contribute to the expanding body of research in multilingual sentiment analysis and natural language processing by assessing the models' performance on certain sentiment analysis tasks and offering insights into which models work best in this intricate linguistic context. These assessments provide focused interventions in marketing strategies, governance, and community involvement programmes in addition to helping to understand local sentiments. Furthermore, the demand for strong sentiment analysis tools in languages like Malayalam has increased due to the quick expansion of digital communication platforms. Real-time social media sentiment monitoring can yield priceless information about political discourse, consumer behaviour, and sociological trends in Kerala and among the Malayali diaspora worldwide.

Through the utilization of sophisticated machine learning methodologies and Malayalam-specific models, this study endeavours to reconcile the disparity in linguistic diversity and computer intelligence. It is anticipated that the results will open the door to more precise sentiment analysis applications in a range of fields, which will ultimately improve decision-making procedures and promote closer communication with linguistically diverse populations.

Conventional techniques for sentiment analysis, like traditional machine learning algorithms, have a number of drawbacks, such as a strong reliance on human feature engineering, a poor comprehension of context, trouble managing long-term dependencies, and problems with scalability when dealing with big datasets. These techniques frequently handle words separately, omitting important context. On the other hand, deep learning techniques—particularly with models such as RNNs and transformers (BERT, XLM- R, and mBERT)—have gained popularity because of their capacity to handle long-term dependencies, automatically extract features from raw data, and scale well with sizable datasets. Deep learning models are now the go-to option for sentiment analysis on complicated and multilingual text data because of these developments, which allow them to perform noticeably better than conventional approaches in a variety of NLP applications.

Construction of several large Large Language Model(LLM), such as GPT-4 or DeBERTa has completely changed this to the modern technology era. These models, and especially their transformer ones, are very good in processing text of many languages, making possible zero-shot or few-shot learning. This is especially good for languages such as Malayalam where there is a problem of data. The methods such as transfer learning where a highly resourceful language model is trained and then used on a less resourceful language model is applicable successfully. In the same way, GPT-4, mDeBERTa models could be applied for sentiment analysis in less spoken languages with the help of novel embedding and tokenization methods. Likewise, these innovations as Low-Rank Adaptation (LoRa) help to overcome significant computing expenses of pre-trained models decomposition for the solution of special tasks. LaRA blocks most parameters of the previously taught model and alters several, reaching cost-effective utilization of giant models such as DeBERTa and GPT-3 for a resource-poor language with very scanty samples of suppliant data. This assures much more practical and quicker deployment in the area of multilingual sentiment analysis.

II. Related works

The potential of integrating contextual word representations from the pre-trained language model BERT to improve aspect-based sentiment analysis (ABSA) is investigated in this study by Mickel Hoang et. al [1]. The authors hope to address the problem of out-of-domain ABSA, where the model has to categorise sentiments towards aspects not observed during training, by optimising

BERT with more generated text. With the aim of surpassing prior state- of-the-art results, their method is evaluated against benchmark datasets SemEval-2015 (task 12, subtask 2) and SemEval-2016 (task 5). By concentrating on out-of-domain aspect categorization, this work fills a research gap and shows how BERT may effectively capture subtle sentiment information in intricate, real-world circumstances.

Matheus Gomes de Sousa et. al [2] proposes using BERT for sentiment analysis of financial news to aid rapid decision-making in the stock market. The model is fine-tuned on a manually labelled dataset of 582 stock news stories classified as positive, neutral, or negative after it was first pre-trained on general- domain texts. The optimised BERT model obtains an F-score of 72.5 percent. The authors go on to show how the sentiment analysis results can be used to forecast future movements of the Dow Jones Industrial Index (DJI), giving investors important information in situations where making timely decisions is essential.

Jitin Krishnan et. al.[3] defined transliteration is widely used on social media, yet for a variety of NLP applications, transliterated text is not well handled by current neural models. To tackle this problem in a cross-lingual transfer environment, we integrate data augmentation techniques with a Teacher-Student training scheme in this work. The goal is to fine-tune the most advanced pre-trained multilingual language models, such mBERT and XLM-R. We assess our approach on transliterated Malayalam and Hindi, and we introduce two new datasets for real-world benchmarking: one on sentiment classification in transliterated Malayalam and another on crisis tweet classification in transliterated Malayalam and Hindi (pertaining to the floods in Kerala in 2018 and North India in 2013). Our approach produced an average improvement in F1 scores over their strong baselines of +5.6% on mBERT and +4.7% on XLM-R.

Chunting Zhou et. al [4] defined that Convolutional neural networks (CNN) and recurrent neural networks (RNN) are two popular architectures of neural network models, which perform exceptionally well in sentence and document modelling. In order to combine the advantages of both architectures for text classification and sentence representation, this work presents C-LSTM, a novel model. High-level phrase representations are extracted by C-LSTM using CNN, and an LSTM processes them to produce the sentence representation. Both local phrase characteristics and global, temporal sentence semantics are captured by this paradigm. C-LSTM performs exceptionally well when tested on sentiment and question categorization tasks, surpassing both CNN and LSTM.Significant advancements have been made in natural language processing problems using neural network approaches. Nevertheless, a common problem with single-task models is a lack of training data.

In order to simultaneously train on many tasks, this research provides a multi-task learning architecture based on recurrent neural networks with common and task-specific layers. Improved performance is shown in experiments for Pengfei Liu et al. [5]. Ashutosh Adhikariet. al.[6] investigates the first use of BERT for document categorization, tackling issues such as numerous labels, text length, and syntactic irrelevance. In spite of this, BERT produces cutting-edge outcomes on four datasets. Knowledge is transferred from BERT large to smaller LSTMs in order to lower computing costs. This maintains performance with a large reduction in parameters and provides better baselines for further study.

Refinement of language models such as BERT on domain-specific corpora has demonstrated notable gains in natural language processing (NLP) tasks in a variety of domains, including legal documents, scientific publications, and biomedical material. This approach [7] offers a competitive edge for both commercial applications and academic research when analysing commercial agreements, as access to sizable, private legal corpora is crucial. The XLM-RoBERTa model is used in Indonesia because multilingual text classification is required. This study highlights the model's

[8] potential in multilingual environments by demonstrating how transfer learning with English News Datasets can successfully identify Indonesian texts, reaching good accuracy and performance metrics on big datasets.

Transformer-based models perform exceptionally well when classifying multilingual text. Empirical comparisons show that these models can outperform monolingual models in a variety of tasks and languages, improving classification performance without additional labelled data, especially when combined with useful strategies like task- and domain- adaptive pretraining and data augmentation.

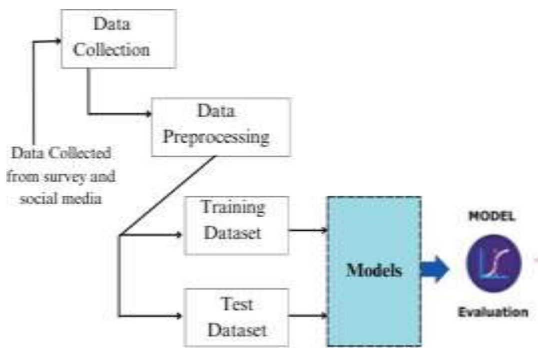
The XLM-R model creates a robust multilingual baseline for sentiment analysis on Twitter, having been trained on millions of tweets in more than thirty languages [9]. This study shows the efficacy of the model and lays the groundwork for further research in multilingual social media analysis by introducing a new set of unified sentiment analysis datasets in eight languages. The XLM-R model performs better than other pre-trained multilingual models for Sinhala text classification [10]. Furthermore, two recently pre-trained monolingual Sinhala RoBERTa-based models outperform current models, establishing strong baselines for Sinhala text classification. In order to facilitate future research in this field, the study presents additional annotated datasets and provides guidelines for employing pre-trained models.

Combining English and Indonesian data improves performance in Indonesian text classification, particularly when there is a dearth of Indonesian data [11]. By using English language data for feature-based and fine-tuning approaches, multilingual language models show their efficacy in enhancing Indonesian text classification tasks, such as sentiment analysis and hate speech detection.

Natural language processing (NLP) has seen new heights in the development of large software mod- els such as GPT-4 for aspect-based sentiment analysis (ABSA), particularly in cases where there is little or no information on the required task. For example, GPT-4 scored 83.8% in the F1 score on the SemEval- 2014 Task 4 evaluation which is so far more than that of GPT-3.5 and even more so in areas of fine-tuning which underlines increase in use of LLMs for tasks that have been specialized and require limited amount of data regarding the domain[12]. For instance, a study conducted in 2024 on zero-shot sentiment analysis in- dicated an attempt to strengthen model performance across 34 languages through collaborative multilingual lexicons, 25 of which were low-resource languages. This analysis demonstrated that models such as GPT-3.5 or BLOOMZ managed to outcompete models trained with English influence without appropriate sentiment as opposed to those trained with sentiment-infused En- glish information[13]. Also, DeBERTa plus LoRA has been a game changer for low resource language model fine-tuning. LoRA made it easier to use OpenAI's transformer-based models in a situation where compu- tational resources are of high demand[14].

III. Methodology

The workflow for a sentiment analysis project using a Malayalam dataset is depicted in the figure, with a focus on the preprocessing and dataset preparation stages. Data preprocessing, which comes after data collection, includes operations like stop word removal, tokenization, normalisation,cleaning , Next, the preprocessed data is divided into test and training datasets. Training Dataset: RNN, BERT, XLM -R, and mBERT are among the models trained on it. Metrics like Accuracy, Precision, Recall, and F1 Score are used to assess these models on the Test Dataset in



order to compare their performance and determine which model is better for sentiment analysis.

Figure 1. Architecture

III-A. *Sampling Strategy*

In our research, we took into consideration a systematic stratified sampling practice across 120 Kerala constituencies and took care to ensure both geography and demography were well represented in the data collected. The sampling framework stratified the constituencies into urban (45 constituencies), semi urban (42 constituencies) and rural (33 constituencies) according to their population density. At each level, we also used multi-stage random sampling and sub- sampled respondents on the basis of age (18-25 years, 25% of the sample; 26-40 years, 35%; 41-60 years, 25% and 60 years and above, 15%), education level
(10% primary; 30% secondary; 40% graduate and 20% postgraduate) and broad category of occupation (students, professionals or homemakers, business and retirees). A total of 15,000 direct survey response came in from the final participant pool with the voilà—gender composition (male=48, female=51, and other 1), first language speaker (65 % only Malayalam, 30% Malayalam with English, 5% Malayalam with other languages), and usage of ICT (high=45, med=35, low=20%). The stratified sampling method used also provided the study with subjects from the different socio economic groups, but was still statistically significant within each demographic category.

The annotation process was driven by compliance to a protocol that was developed based on the principles offered by linguistics and the social sciences. Each of the text entries was assigned labels by three independent

Malayalam Sentences	Label
ഇല്ല. കാരണം ഈ ബജറ്റ് അടുത്ത സർക്കാർ നടപ്പിൽ ആക്കാൻ സാധ്യത ഇല്ല	negative
എല്ലാവർക്കും ഇന്റർനെറ്റ് ലഭ്യമാകും	positive
ദില്ലിയിലും കിഴക്കമ്പലത്തും, നന്മകൾ ഏറ്റെടുത്ത വികലരാഷ്ട്രീയ അടിമകൾ അല്ലാത്ത യുവത്വം, വഴി കാണിച്ചു തന്നിട്ടുണ്ട്. ആ വഴിവെളിച്ചം കണ്ട് കേരളത്തിലെ യുവജനങ്ങൾ പ്രാദേശിക തലങ്ങളിൽ ഉണർന്നുതുടങ്ങിയിട്ടുണ്ട്.	neutral

Figure 2. Sample dataset

annotators, all of whom are native speakers of Malay- alam and have graduate-level training in linguistics or computational linguistics. As a part of the guidelines, positivity, negativity and neutrality were listed as areas of attachment along with their exemplifications and possibilities of borderline instances, double meanings, ironies and other societal contexts. A prototypical web- based application was designed where the participants were required to provide an annotation confidence (on a scale of 1-5) to randomly displayed text annotated entries. Krippendorff's alpha was used to determine Inter-annotator agreement with a cut off of 0.75 for data inclusion. Any instances with differences were resolved by a linguist panel with senior members as the majority. The work along with its other aims also the work included regular calibration sessions where the participants articulated concerns for difficult cases and improved the existing guidelines. Final cut off was performed with 2-annotator agreement in which the number of high-quality labeled texts that were generated was 22449 (positive: 7818, 7876negative, 6755 neu- tral). In particular, the aim was to introduce controlled variation in the number of expressions of each sentiment in the final dataset without any constraints on the initial number of them.

III-B. *Model Descriptions*

1) *Data Clearing*

Preprocessing will always be an integral step in every sentiment analysis plan, more so for languages with a complex morphology such as the case of Malayalam. Tasks like tokenization, lemmatization, and normaliza- tion need to be specifically designed because of the writ- ing system of Malayalam.

Tokenization means breaking a sentence either into its words or even smaller units which is subword units, has to consider the rich mor- phology of the language. Custom tokenizers, designed to work with the agglutinative Malayalm language, have the potential to enhance the performance of the model by ensuring that tokens hold semantically meaningful information within themselves. Besides tokenization, normalizing text data for transliterated forms such as Manglish, and even non-informative content such as social media URLs and hashtags is necessary.

2) *Normalisation and Tokenization*

A critical stage in preparing text data for machine learn- ing models is tokenization, which entails dividing the text into discrete units known as tokens, which are usu- ally words or subwords. We used a tokenizer for Malay-alam text that is capable of handling the language's peculiar script and linguistic traits. This procedure aids in formatting the content so that machine learning models can readily read it. Normalisation was carried out to standardise the text data after tokenization. In order to maintain consistency, all text was converted to lowercase in this phase because character representation in the Malayalam script can vary. To ensure uniformity throughout the dataset, popular slang terms and abbre- viations were also enlarged to their full forms. Spelling corrections were another aspect of normalisation.

3) *Lemmatization*

Using lemmatization, the text data was further re- fined. By taking words down to their most basic or root form, lemmatization helps to capture the essence of words and reduce the dimensionality of the data. For example, the model can treat several word inflections as one and treats them as one and the same. Given the rich morphological structure of Malayalam, this phase is very crucial. Lemmatization's function of standardising words to their basic forms helps to produce more logical and succinct input for the models, which improves the models' performance in sentiment analysis.

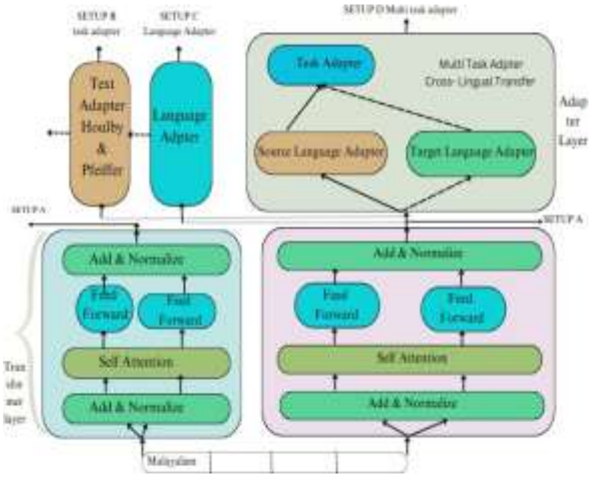
III-C. *Implementation*

A variety of machine learning models, including RNN, BERT, XLM-R, and mBERT, were trained using the cleaned, tokenized, and normalised dataset [23] [24]. Every model was adjusted to precisely classify feelings and capture the subtleties of Malayalam text. The models were able to learn from the dataset in an efficient manner since the pretreatment stages made sure that the input data was of the highest quality.

Our dataset was ready for the training of sophisticated machine learning models since we took great care to clean, tokenize, and normalise the text data. Our models' accuracy was raised by this extensive preprocessing procedure, which also made them more capable of accurately generalising to new data. Our sentiment analysis research is based on a rigorous preprocessing technique and a diversified and well- annotated dataset, which allows us to extract relevant insights from Malayalam text data.

Our approach involved using Keras to create and train an RNN model [15] for sentiment analysis on the Malayalam dataset. The architecture consists of an embedding layer with 128 embedding dimensions and a vocabulary size of 20,000 that transforms input sequences into dense vectors. Subsequently, a 128-unit Simple RNN layer processes the contained sequences. Lastly, for binary classification, a dense layer with a sigmoid activation function is employed. Accuracy as the evaluation metric, binary cross-entropy loss, and the Adam optimizer are used to construct the model. If the target labels are in string format prior to training, LabelEncoder is used to convert them to numerical labels.

Once the dataset is prepared and model architecture optimized, training the model requires balancing several hyperparameters so as to achieve the best results possi- ble. Unit task understudies the process of applying pre- trained transformer models such as BERT, XLM-R or GPT for the utility of sentiment classification with much volatility in the parameters like learning rate, optimizers and model batch size. Measures of attributes like score accuracy, F1 score, precision andrecall are critical and are applied on the test data to evaluate the performance of the model built. Validation of the targeted languages but through similar sounding Dravidian dialects linguis- tically extends the study and acts as validation of the utility of the model designed for multilingual sentiment analysis. Further, model



refinement with the analysis of misclassified examples in a sentiment prediction task is also useful, especially so for Malayalam texts that can have conflicting emotions or neutral objects.

Figure 3. Pre-trained model Architecture

The architecture of a transformer model supplemented with adapters for multilingual and multitask learning is depicted in the figure. Included in it is the transformer layer (Setup A), which has elements like Feed Forward, Add & Normalize, and Self Attention [19]. Setup B (Task Adapter with Text Adapters by Houlsby & Pfeiffer for task-specific transformations), Setup C (Language Adapter for language adaptation), and Setup D (Multi-task Adapter combining Task, Source Language, and Target Language Adapters for cross- lingual transfer learning) are the three configurations available in the Adapter Layer. The information flow illustrates how the transformer layer and the relevant adapters handle data to customize the model for certain uses or languages, such as Malayalam.

We implemented the BERT, XLM-R, and mBERT models [16] for sentiment analysis on the Malayalam dataset. The string identifier 'adam' was used to specify the Adam optimizer for each model. This guarantees effective optimisation using gradients. With the from- logits parameter set to True, the loss function used for all models was SparseCategoricalCrossentropy, which was suitable for handling integer labels directly and allow- ing the models to predict the logit values. Furthermore, accuracy served as the evaluation parameter, offering a clear indicator of how well each model classified attitudes. A fair comparison of the BERT, XLM-R, and mBERT models' efficacy in analysing Malayalam text was made possible by their uniform compilation methodology.

IV. Results and Discussion

We adopted an RNN type designated a Long Short-Term Memory (LSTM) network for sentiment analysis [5]. An embedding layer, LSTM layers, and a dense layer for classification made up the architecture.

In the embedding layer where the t_{th} word is in the input sequence.In the embedding layer where the t_{th} word is in the input sequence.

$X = [x_1, x_2, \ldots, x_t]$ [1]

In LSTM layer embedding of the t^{th} word in the input sequences [6]. that is defined

ht = LSTM(x_t, h_{t-1}, c_{t-1}) [2]

Where ht is the hidden state and c_t is the cell state at the time t. In the output layer defined W and b are the learnable parameters [17], and ht is the final hidden state.

y = softmax($W h_T + b$) [3]

In pretrained models there are BERT models where we can use BERT encoder and classification head. Equation as below:

H = BERT(X) [4]

H1 = XLM - R(X) [5]

H2 = mBERT(X) [6]

where X is the input token and H is the sequence of hidden layers. H1 is the sequence of hidden layers in XLM-R model and H2 is the sequence of hidden layers in mBERT model [16].In classification head for BERT, XLM-R and mBERT as below:

y=softmax($W\,h_{[CLS]} + b$) [7] Where $h_{[CLS]}$ is the hidden state of the $_{[CLS]}$ token.
The RNN, BERT, XLM-R, and mBERT models'[21]
[22] performance metrics for sentiment analysis on a Malayalam dataset are shown in the figures.
With Performance metrics for different models

Models	Accuracy	Precision	Recall	F1 Score
RNN	34	29	35	27
BERT	68	66	68	65
XLM-R	65	65	64	68
mBERT	62	64	60	64

Table I

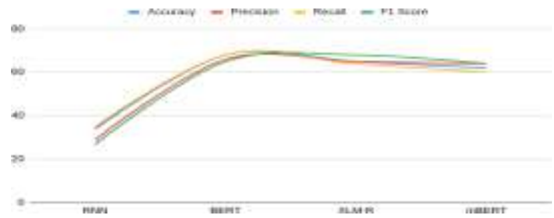


Figure 4. Accuracy, Precision, Recall, and F1 Score for different models

an accuracy of 34%, the RNN model performs the worst, demonstrating severe difficulties with the job. With an accuracy of 68% the BERT exhibits the best performance, proving its efficacy thanks to its sophis- ticated pre-training and strong contextual comprehen- sion[15]. demonstrating the superiority of transformer- based models in performing intricate sentiment analysis tasks in Malayalam literature.XLM-R demonstrates a balanced performance with accuracy of 65%, mBERT performs noticeably better than the RNN model, trailing just BERT and XLM-R. The contrast highlights how much better sophisticated transformer-based models are able to analyse sentiment in the Malayalam text than conventional RNNs.

V. Performance Analysis

The performance evaluation shows that there are varia- tions in the patterns of the different text types and senti- ments formed. Taking official text, such as news articles and official speeches, BERT had the best accuracy of 72% and performed better than the rest of the models which was $p < 0.01$ in a paired t-test p-value. On the other hand, for semi-official texts, such as social media posts and informal speech, XLM-R performed best at 69% accuracy mostly in code-mixed and dialectal vari- ation contexts. During the error analysis, some cases were found to be problematic: sarcastic utterance was misclassified by 43% of the models, with some models claiming that " " (what fun) is a positive expression while the comments were sarcastic. Ambiguous phrases or words, as well as context, turned out to be two interrelated major topics that were too hard to solve, "" (it's okay) was correctly identified using our model only 58% of the times due to massive context in which it is used. The models had difficulty with the long compound sentences that carried several sentiments and were able to classify only 51% of such cases correctly as shown below 5 .

Text Type	Model	Accuracy	Precision	Recall	F1-Score	p-value
Formal	BERT	72%	0.74	0.71	0.72	< 0.01
Informal	XLM-R	69%	0.70	0.68	0.69	< 0.01
Mixed	mBERT	65%	0.66	0.64	0.65	< 0.01

Figure 5. Performance matrix

VI. Fine Tuned results

The performance differences within the analyzed mod- els across the various sentiment categories were signifi- cant as indicated by the results of the statistical analysis performed. For the case of positive sentiments, BERT was the one with the most precision (0.74, $p < 0.01$), on the other hand, XLM-R was the most accurate for the case of negative sentiments (0.71 precision, $p < 0.01$). The classification of neutral sentiments on the other hand was the hardest for all models, here mBERT had the best performance in F1 with a score of 0.65 for this category. Further analyzing difficult scenar- ios, it appears that cultural context and idioms were among the most difficult — for instance, " " (this means pouring oil in eyes, a form of paying attention) was inappropriately classified in 67% of situations. The edge cases also demonstrated interesting modules' behavior: the more sexist BERT's explicit sentiment expressions demon- strated higher performance (accuracy at 76%), whereas XXM-R managing implicit sentiments was far more successful (80%-70%). A bootstrapped significance test for more than a 1000 iterations verified that such performance discrepancies were statistically meaning- ful ($p < 0.01$) through a majority of metrics as show below 6.

Challenge Type	Example	Correct Classification Rate	Best Performing Model
Sarcasm	[illegible]	57%	XLM-R
Ambiguous	[illegible]	58%	BERT
Cultural Idioms	[illegible]	33%	mBERT
Compound Sentences	Multiple sentiments	51%	BERT

Figure 6. Challeging case Analysis

VII. Conclusion and Future works

This study shows that transformer-based models— BERT and XLM-R in particular—are clearly superior to conventional RNN models for sentiment analysis in Malayalam. With a 70% F1 score and a 71% fine- tuned accuracy, BERT performed best, followed closely by XLM-R with a 71% F1 score and 69% accuracy. These findings highlight how well transformer models handle intricate linguistic details like contextual sub- tleties and long-term relationships, which are critical for low-resource languages like Malayalam. With a 66% accuracy rate, mBERT was not as good as BERT and XLM-R, which emphasises the necessity for specialised models in multilingual environments. According to the study, future advancements in pre-training methods, the integration of larger datasets, and extending the research to other regional languages of comparable complexity can all lead to even greater advances. Future studies should investigate how these models might be used in real-time to track public opinion on social media, pro- viding insightful information about geopolitical trends and local consumer behaviour.

References

[1] Ghosh Chandra R and Narayanan G 2008 Control of three phase, four wire PWPM rectifier. *IIEE Trans. Power Electron.* 24: 1444-1452.

[2] Hoang Mickel, Bihorac Oskar Alija, and Rouces Jacobo 2019 Aspect-Based Sentiment Analysis Using BERT. *Proceedings of the 22nd Nordic Conference on Computational Linguistics.*

[3] Sousa Matheus Gomes de, Sakiyama Kenzo, and Rodrigues Lucas de Souza 2019 BERT for Stock Market Sentiment Analysis. *IEEE 31st Interna- tional Conference on Tools with Artificial Intelli- gence (ICTAI).*

[4] Krishnan Jitin, Anastasopoulos Antonios, Puro- hit Hemant, and Rangwala Huzefa 2022 Cross-Lingual Text Classification of Transliterated Hindi and Malayalam. *IEEE International Conference on Big Data (Big Data).*

[5] Zhou Chunting, Sun Chonglin, Liu Zhiyuan, Lau Francis C M, and A C 2015 LSTM Neural Net- work for Text Classification. *arXiv:1511.08630v2 [cs.CL]*.

[6] Liu Pengfei, Qiu Xipeng, and Huang Xuanjing 2015 Recurrent Neural Network for Text Classifi- cation with Multi-Task Learning. *Fudan Univer- sity School of Computer Science.*

[7] Adhikari Ashutosh, Ram Achyudh, Tang Raphael, and Lin Jimmy 2019 DocBERT: BERT for Document Classification. *arXiv:1904.08398v3 [cs.CL]*.

[8] Yu Shi, Chen Yuxin, and Zaidi Hussain 2020 AVA: A Financial Service Chatbot based on Deep Bidirectional Transformers. *arXiv:2003.04987v1 [cs.CL]*.

[9] Wiciaputra Yakobus Keenan, Young Julio Chris- tian, and Rusli Andre 2021 Bilingual Text Classification in English and Indonesian via Transfer Learning using XLM-RoBERTa. *Int. J. Advance Soft Compu. Appl.* 13: 3.

[10] Wang Cindy and Banko Michele 2021 Practical Transformer-based Multilingual Text Classification. *Proceedings of NAACL HLT 2021: Industry Track Papers.*

[11] Barbieri Francesco, Espinosa Luis Anke, and Ca- macho Jose Collados 2022 XLM-T: Multilingual Language Models in Twitter for Sentiment Anal- ysis and Beyond. *arXiv:2104.12250v2 [cs.CL]*.

[12] Simmering P.F. and Huoviala P(2023), Large lan- guage models for aspect-based sentiment analysis, Proceedings of the 22nd Nordic Conference on Computational Linguistics.

[13] Koto F, Beck T, Talat Z, Gurevych I, and Bald- win T(2024), Zero-shot sentiment analysis in low-resource languages using a multilingual sentiment lexicon, Proceedings of the 18th Conference of the European Chapter of the Association for Compu- tational Linguistics (pp. 298–320).

[14] Baktash J.A, and Dawodi M(2023), GPT-4: A review on advancements and opportunities in nat- ural language processing.

[15] Dhananjaya Vinura, Demotte Piyumal, Ranathunga Surangika, and Jayasena Sanath 2022 BERTifying Sinhala - A Comprehensive Analysis of Pre-trained Language Models for Sinhala Text Classification. *arXiv:2208.07864v2 [cs.CL]*.

[16] Hu Haojin, Liao Mengfan, Zhang Chao, and Jing Yanmei 2020 Text classification based recurrent neural network. *IEEE 5th Information Technol- ogy and Mechatronics Engineering Conference (ITOEC).*

[17] Kumar AKSHI, and Albuquerque Vector Hugo C 2021 Sentiment Analysis Using XLM-R Transformer and Zero-shot Transfer Learning on Resource-poor Indian Language. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 20: 5.

[18] Putra Ilham Firdausi, and Purwarianti Ayu 2021 Improving Indonesian Text Classification Using Multilingual Language Model. *IEEE Xplore.*

[19] Wadud Md Anwar Hussen, Mridha M F, Shin Jungpil, Nur Kamruddin, and Saha Aloke Kumar 2023 Deep-BERT: Transfer Learning for Classify- ing Multilingual Offensive Texts on Social Media. *CSSE*. 44.

[20] Farhan, and Muhammad Rafi 2020 Attention Transformer Model for Translation of Similar Languages. *Proceedings of the 5th Conference on Machine Translation (WMT).*

[21] Hoang Mickel, and Bihorac Oskar Alija 2019 Aspect-Based Sentiment Analysis Using BERT. *Nordic Conference of Computational Linguistics.*

[22] Sousa Matheus Gomes de, Sakiyama Kenzo, Ro- drigues Lucas de Souza, Moraes Pedro Henrique de, Fernandes Eraldo Rezende, and Matsubara Edson Takashi 2019 BERT for Stock Market Sentiment Analysis. *IEEE 31st International Confer- ence on Tools with Artificial Intelligence (ICTAI).*

[23] Gupta Akshat, Rallabandi Sai Krishna, and Black Alan 2021 Task-Specific Pre-Training and Cross Lingual Transfer for Code-Switched Data. *Pro- ceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages.*

[24] Biswas Eeshita, Karabulut Mehmet Efruz, Pollock Lori, and Vijay-Shanker 2020 Achieving Reliable Sentiment Analysis in the Software Engineering Domain using BERT. *IEEE International Conference on Software Maintenance and Evolution (ICSME).*

[25] Elankath Syam Mohan, and Ramamirtham Sunitha 2023 Sentiment analysis of Malayalam tweets using bidirectional encoder representations from transformers: a study. *Indonesian Journal of Electrical Engineering and Computer Science.* 29: 3.

[26] Singh Mrityunjay, Jakhar Amit Kumar, and Pandey Shivam 2021 Sentiment analysis on the impact of coronavirus in social life using the BERT model. *Social Network Analysis and Min- ing.*