Sentiment Analysis of Sinhala News Comments Using Transformers

Isuru Bandaranayake and Hakim Usoof
Department of Statistics & Computer Science
Faculty of Science, University of Peradeniya
Peradeniya, Sri Lanka
bandaranayakeisuru@gmail.com, hau@sci.pdn.ac.lk

Abstract

Sentiment analysis has witnessed significant advancements with the emergence of deep learning models such as transformer models. Transformer models adopt the mechanism of self-attention and have achieved state-of-the-art performance across various natural language processing (NLP) tasks, including sentiment analysis. However, limited studies are exploring the application of these recent advancements in sentiment analysis of Sinhala text. This study addresses this research gap by employing transformer models such as BERT, DistilBERT, RoBERTa, and XLM-RoBERTa (XLM-R) for sentiment analysis of Sinhala news comments. This study was conducted for 4 classes: positive, negative, neutral, and conflict, as well as for 3 classes: positive, negative, and neutral. It revealed that the XLM-R-large model outperformed the other four models, and the transformer models used in previous studies for the Sinhala language. The XLM-R-large model achieved an accuracy of 65.84% and a macro-F1 score of 62.04% for sentiment analysis with four classes and an accuracy of 75.90% and a macro-F1 score of 72.31% for three classes.

1 Introduction

Sentiment analysis is a fundamental task in NLP which aims to analyze and understand the sentiment expressed in textual data. While sentiment analysis has been extensively studied for major languages such as English, research on low-resource languages is relatively limited.

Sinhala, a morphologically rich Indo-Aryan language, serves as the native language of the Sinhalese people, constituting a significant portion of the population in Sri Lanka with an estimated count of 20 million speakers. However, despite its large speaker base, Sinhala is considered a low-resource language in the context of NLP research due to the scarcity of available linguistic resources for analysis and processing (de Silva, 2019).

Sentiment analysis has experienced significant progress with the advent of large-scale pre-trained language models (Mishev et al., 2020). These models have demonstrated promising results in text classification tasks for high-resource and low-resource languages. Transformer models have revolutionized NLP tasks by leveraging attention mechanisms and self-attention layers, allowing them to capture intricate linguistic patterns and dependencies (Devlin et al., 2018). Notably, transformer-based models such as BERT (Devlin et al., 2018), RoBERTa (Liu et al., 2019), and XLM-R (Conneau et al., 2019) have shown remarkable performance across various languages, making them promising candidates for sentiment analysis in Sinhala.

One of the primary advantages of employing transformer models for sentiment analysis in Sinhala is their ability to handle the language's morphological richness and syntactic complexities. Sinhala exhibits complicated morphological variations and context-dependent sentiment expressions (Medagoda, 2017), which transformer models can effectively capture.

However, applying transformer models to sentiment analysis in Sinhala also poses specific challenges. One major challenge is the scarcity of

Proceedings of the First Workshop on Natural Language Processing for Indo-Aryan and Dravidian Languages (IndoNLP2025), pages 99–107
January 20, 2025. ©2025 Association for Computational Linguistics

annotated sentiment datasets for fine-tuning transformer models. There exists a sentiment dataset of 15,059 Sinhala news comments, annotated with four classes: Positive, Negative, Neutral, and Conflict (Senevirathne et al., 2020). However, the limited size of this dataset hinders the ability of transformer models to achieve optimal performance.

To address this limitation, we expanded the existing Sinhala news comments dataset by adding 5,000 annotated comments to the dataset. While the dataset size may still be considered limited, this extension introduced more diverse examples and enabled some level of expansion for training and evaluation purposes.

In this research, we conducted two sentiment analysis experiments considering four sentiment classes and three sentiment classes respectively. The goal was to evaluate the performance of monolingual models such as BERT, DistilBERT, and RoBERTa as well as multilingual models such as XLM-R-base and XLM-R-large models in sentiment analysis for the Sinhala language. We investigated their capabilities in effectively capturing sentiment information, accommodating the morphological variations of the language, and addressing the limited availability of labeled data. These research outcomes will contribute valuable insights to the field of sentiment analysis in Sinhala and will provide a foundation for future studies and applications.

2 Related Work

Recent developments in deep learning techniques have made it possible to achieve better results in the domain of NLP. Deep learning techniques do not use language-dependent features. Therefore, deep learning techniques have outperformed traditional statistical machine learning techniques (dos Santos and Gatti, 2014). Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) were the most popular deep learning techniques used in the NLP domain until Long Short-Term Memory (LSTM) and Transformer models were introduced. Kim (2014) proposed a method using CNN with hyperparameter tuning for sentiment analysis, and it was shown that a simple CNN with one layer of convolution and little hyperparameter tuning performs remarkably well. LSTM encoders were experimented for sentiment analysis by Yang et al. (2016) and bi-directional LSTM by G. Xu et al. (2019). Both studies showed improved results compared to previous studies, which used deep learning techniques such as CNN and RNN. An attention-based Bi-LSTM with a convolutional layer scheme called AC-BiLSTM was proposed by W. Liu et al. (2017) for sentiment analysis. Word2Vec, which is one of the most popular word-embedding models, was introduced by Goldberg & Levy (2014). Word2Vec improved the efficiency of the training procedure and enhanced the training speed and accuracy. An improved version of the Word2Vec model called GloVe was introduced by Pennington et al. (2014). GloVe outperformed other models on word analogy, word similarity, and named entity recognition tasks. Transformer models were introduced by Vaswani et al. (2017). Transformers could train significantly faster than architectures based on recurrent or convolutional layers. H. Xu et al. (2019) carried out aspect-based sentiment analysis using the BERT model, producing a state-of-the-art performance for sentiment analysis. Liao et al. (2021) used RoBERTa, an improved version of BERT, to carry out aspect-category sentiment analysis and it outperformed other models for comparison in aspect-category sentiment analysis.

Since Sinhala is a low-resource language, research done on the Sinhala language is very limited. The first sentiment analysis for the Sinhala language was carried out by N. Medagoda et al. (2015) by constructing a sentiment lexicon for Sinhala with the aid of the SentiWordNet 3.0, an English sentiment lexicon. It achieved a maximum accuracy of 60% in Naïve Bayes (NB) classification. The first sentiment analysis for the Sinhala language using an artificial neural network was conducted by N. Medagoda (2016) using a simple feed-forward neural network and part of speech tags as a feature. This model achieved an accuracy of 55%. Chathuranga et al. (2019) used a rule-based technique for binary sentiment classification of Sinhala news comments. In this study, they generated a Sinhala sentiment lexicon in a semi-automated way and used it for sentiment classification of Sinhala news comments. NB, Support Vector Machines (SVM), and decision trees were used in this study and obtained accuracy between 65% - 70%. The best accuracy of 69.23% was obtained for the NB model. Ranathunga & Liyanage (2021) conducted sentiment analysis for Sinhala news comments with deep learning techniques such as LSTM and CNN+SVM. Also, this study experimented with Word2Vec and

fastText word embeddings for Sinhala (Ranathunga and Liyanage, 2021). Further, statistical machine learning algorithms such as NB, logistic regression, decision trees, random forests, and SVM were experimented by training them with the same features and conducting a sentiment analysis for Sinhala news comments. This research was carried out to study the use of various models with respect to the dimensionality of the embeddings and the effect of punctuation marks (Ranathunga and Liyanage, 2021). Demotte et al. (2020) used an approach based on the S-LSTM model for sentiment analysis of Sinhala news comments. The same dataset used by Ranathunga & Liyanage (2021) was used in this study, and it was found that S-LSTM outperforms the traditional LSTM used in the study conducted by Ranathunga & Liyanage (2021). Senevirathne et al. (2020) conducted comprehensive research on the use of RNN, LSTM, and Bi-LSTM models as well as more recent models such as hierarchical attention hybrid neural networks and capsule networks for sentiment analysis. As part of this study, they released a dataset of 15059 Sinhala news comments, annotated with four classes (Positive, Negative, Neutral, and Conflict) and a corpus of 9.48 million tokens (Senevirathne et al., 2020). Dhananjaya et al. (2022) conducted experiments to explore the performance of transformer models in various linguistic tasks, including sentiment analysis, for the Sinhala language. Their study evaluated LASER, LaBSE, XLM-R-large, XLM-R-base, and three RoBERTa-based models pre-trained specifically for Sinhala: SinBERT, SinBERTo, and SinhalaBERTo.

3 Models

In this study, we used the following transformer models to carry out sentiment analysis for the Sinhala language,

- BERT: Bidirectional Encoder Representations from Transformers
- DistilBERT: Distilled version of BERT
- RoBERTa: Robustly Optimized BERT Pretraining Approach
- XLM-R: Cross-lingual Language Model – RoBERTa
    - XLM-R-base
    - XLM-R-large

BERT, which stands for Bi-directional Encoder Representations from Transformers, is a bidirectional transformer model pre-trained on Toronto Book Corpus and Wikipedia. BERT was developed by Google, and it was the state-of-the-art language model for NLP tasks at the time it was released (Devlin et al., 2018).

DistilBERT is a lighter and faster version of the BERT model, and it was developed by Huggingface. DistilBERT has the same general architecture as BERT, but the size is 40% less than that of BERT and retains 97% of the language understanding capabilities of BERT. Also, DistilBERT is 60% faster than BERT, which is another benefit of this model (Sanh et al., 2019).

RoBERTa stands for Robustly Optimized BERT Pre-training Approach. It is an improved version of the BERT model. RoBERTa has the same architecture as the BERT model but is trained with more data and has better parameter settings (Liu et al., 2019).

XLM-R is a multilingual model pre-trained on filtered Common Crawl data containing more than 100 languages, including Sinhala. This model was developed and released by Facebook AI in 2019 (Conneau et al., 2019). XLM-R model outperformed the multilingual BERT (mBERT) and achieved state-of-the-art results on multiple cross-lingual benchmarks (Conneau et al., 2019). This model can be directly fine-tuned for a downstream task without pre-training on a Sinhala corpus, as this model is already pre-trained on Sinhala. XLM-R consists of two variants: XLM-R-base and XLM-R-large. XLM-R-base is the base version with fewer parameters.

4 Dataset

This study required two datasets to carry out pre-training and fine-tuning of the models. Since the pre-training is unsupervised, it does not require a labeled dataset. However, it required two separate datasets annotated with four classes (Positive, Negative, Neutral, and Conflict) and three classes (Positive, Negative, and Neutral) to fine-tune the models.

4.1 Dataset for pre-training

We used the Sinhala corpus extracted from "Open Super-large Crawled Aggregated coRpus" (OSCAR) dataset to pre-train the models. OSCAR dataset is a multilingual corpus obtained by language classification and filtering of the Common Crawl corpus using the Ungoliant architecture. Common Crawl corpus is a huge

corpus that contains petabytes of raw web page data, metadata extracts, and text extracts gathered over 12 years of web crawling (Abadji et al., 2022). The OSCAR dataset has raw text from 162 languages, including the Sinhala language. This dataset contains 108,593 documents in the Sinhala language and 113,179,741 Sinhala words. The total size of the dataset is around 2.0 GB (Abadji et al., 2022).

4.2 Dataset for fine-tuning

The dataset [1] published by Senevirathne et al. (2020) contains 15059 news comments annotated with four classes: Positive, Negative, Neutral, and Conflict. This dataset contains 9,059 news comments extracted from the Lankadeepa online newspaper [2] by Ranathunga & Liyanage (2021), along with 6,000 news comments extracted from the GossipLanka news website[3]. This annotation has been done by three annotators following the guidelines mentioned below,

- A comment is annotated as positive or negative if it expresses purely a positive or negative opinion.
- A comment is annotated as a conflict if it gives both positive and negative opinions.
- A comment is annotated as neutral if it does not give any positive or negative opinion.

In this study, we expanded this dataset by following the steps below.

Data collection: We collected 803,623 news comments from the GossipLanka news website and filtered them to include only comments written in Sinhala Unicode characters. These comments were then cleaned by removing any characters outside the Unicode range (0D80 - 0DFF). The final dataset of Sinhala news comments contained 417,332 comments.

Classes	Dataset 1 (Four Classes)	Dataset 2 (Three Classes)
Positive	3,587	4,414
Negative	10,228	11,639
Neutral	3,822	3,822
Conflict	2,238	0
Total	19,875	19,875

Table 1: Distribution of comments per class

Data annotation: Two annotators who are native Sinhala speakers carried out the annotating task following the guidelines mentioned previously. We used Cohen's Kappa measure to evaluate the inter-annotator agreement, which yielded a value of 0.794. Both annotators collectively annotated 5,037 Sinhala news comments with four classes (Positive, Negative, Neutral, and Conflict). These annotated comments were added to the existing Sinhala news comments dataset. We carefully removed duplicate entries from the combined dataset to ensure data integrity. The final dataset comprised 19,875 unique comments.

The newly generated dataset was annotated again using Positive, Negative, and Neutral to create a sentiment dataset with three classes. Comments initially labeled as Conflict were annotated as Positive or Negative based on their predominant sentiment. Table 1 shows the distribution of comments per class in the two datasets.

4.3 Model pre-training

Pre-training the XLM-R-base and XLM-R-large models for Sinhala was not required, as these models are already pre-trained on a multilingual corpus that includes Sinhala. However, we had to pre-train the other three models for the Sinhala language, and these were pre-trained using the Sinhala dataset extracted from the OSCAR dataset.

BERT	DistilBERT	RoBERTa
[PAD]	[PAD]	<s>
[UNK]	[UNK]	<pad>
[CLS]	[CLS]	</s>
[SEP]	[SEP]	<unk>
[MASK]	[MASK]	<mask>

Table 2: Special tokens included in tokenizers

Since models cannot process raw data directly, they need to be converted to a representation that the models can process. Therefore, it was necessary to train tokenizers for these models from scratch. BERT and DistilBERT tokenizers use the WordPiece method (Devlin et al., 2018; Sanh et al., 2019), while the RoBERTa tokenizer uses the Byte-

[1] https://github.com/LahiruSen/sinhala_sentiment_anlaysis_tallip

[2] https://www.lankadeepa.lk/

[3] https://www.gossiplankanews.com/

Models	Four Classes				Three Classes			
	F1	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall
BERT	46.34%	47.07%	48.12%	52.13%	59.19%	63.32%	58.22%	61.36%
DistilBERT	49.49%	51.72%	49.59%	53.69%	60.63%	65.13%	60.59%	61.66%
RoBERTa	37.50%	37.36%	41.08%	46.27%	53.17%	56.48%	53.08%	56.67%
XLM-R_{base}	59.16%	62.84%	59.17%	61.85%	69.52%	73.56%	68.45%	71.06%
XLM-R_{large}	62.04%	65.84%	61.79%	64.48%	72.31%	75.90%	72.02%	73.20%

Table 3: Results for sentiment analysis using four classes and three classes

Pair Encoding method (Liu et al., 2019). Tokenizers for the three models were trained with a vocabulary size of 52,000 and a minimum frequency of 2 using the Sinhala dataset extracted from the OSCAR dataset. The vocabulary size defines the number of tokens and alphabets included in the final vocabulary, and the minimum frequency defines the minimum frequency a pair should have to be merged. Special tokens included in BERT, DistilBERT, and RoBERTa tokenizers are listed in Table 2.

After training the tokenizers, the three models were trained for masked language modeling task using the same dataset that was used to train the tokenizers. The models were trained with a vocabulary size of 52,000, a maximum position embedding of 512, a hidden size of 768, 12 attention heads, and 12 hidden layers. During training, tokens in the input sequences were randomly masked with a probability of 0.15. This means that, for each input sequence, approximately 15% of the tokens were selected at random to be replaced with a special token: [MASK] for BERT and DistilBERT, and <mask> for RoBERTa. We used AdamW as an optimizer with a learning rate of 5×10^{-5} and a batch size of 16.

The training of both the models and tokenizers was conducted in the Google Colaboratory environment with V100 16GB GPUs. Due to the high computational cost of pre-training, the models were trained for only one epoch.

4.4 Model fine-tuning

The pre-trained models should be fine-tuned to carry out sentiment analysis. Even though XLM-R-base and XLM-R-large models are already pre-trained for the Sinhala language, it needs to be fine-tuned for sentiment analysis in the Sinhala language. Therefore, all five pre-trained models were fine-tuned for sentiment analysis. Each pre-trained model was fine-tuned twice using Dataset 1 and Dataset 2 separately. According to the original paper of BERT, the recommended number of epochs for fine-tuning a model is 2, 3 and 4 (Devlin et al., 2018). Therefore, BERT and DistilBERT models were fine-tuned for five epochs and at the end of each epoch, the trained model was saved as a checkpoint. The best performing model was selected from the saved checkpoints by considering the loss at each epoch. Similarly, the other three models were fine-tuned for five epochs, and the best performing checkpoint was chosen.

The fine-tuning process for the models involved the use of a consistent set of parameters across BERT, DistilBERT, RoBERTa, XLM-R-base, and XLM-R-large models. For all models, the batch size was set to 16, and a dropout rate of 0.1 was applied to prevent overfitting.

The learning rates were adjusted to optimize training performance, with BERT and DistilBERT using a rate of 2×10^{-5}, RoBERTa using 1×10^{-5}, and both XLM-R-base and XLM-R-large using a rate of 5×10^{-6}. Weight decay was uniformly applied at 0.01 for all models to control overfitting further. The training was conducted using the AdamW optimizer to ensure stable convergence.

5 Results and Discussion

We evaluated the performance of the fine-tuned models using accuracy, macro-F1 score, macro-precision, and macro-recall. The results obtained by Dhananjaya et al. (2022) for the sentiment task serve as the baseline for our study. Table 3 presents the results obtained for sentiment analysis for three and four classes. In this study, we conducted all model training and evaluation using the Transformers library provided by HuggingFace (Wolf et al., 2019) on the Google Colaboratory environment.

For sentiment analysis using four classes, we observe that XLM-R-large achieved the highest macro-F1 score of 62.04%, followed closely by XLM-R-base with a macro-F1 score of 59.16%. Similarly, XLM-R-large continues to display

superior performance for sentiment analysis using three classes, achieving a macro F1-score of 72.31% and an accuracy of 75.90%. Dhananjaya et al. (2022) obtained a macro-F1 score of 60.45% for sentiment analysis using four classes, which serve as the baseline model. This indicates that our model outperformed the baseline model slightly. One potential reason for the improved performance of XLM-R-large is the utilization of a larger training dataset, allowing the model to learn from a more diverse set of examples and generalize better.

Our study observed that BERT and DistilBERT achieved competitive macro-F1 scores and accuracy for both sentiment analysis tasks with four and three classes. However, the macro-F1 scores of BERT, DistilBERT, and RoBERTa were relatively lower than XLM-R models. The outcome of these monolingual models achieving lower results than XLM-R models was unexpected. Monolingual models are typically trained specifically for a single language, and they would have a better understanding of linguistic patterns, leading to better performance in sentiment analysis tasks. However, the observed results highlighted that XLM-R models performed better in sentiment analysis for Sinhala despite being pre-trained on a multilingual corpus. The reason for this unexpected outcome is the difference in the pre-training process. BERT, DistilBERT, and RoBERTa models were pre-trained for only one epoch, while XLM-R models were pre-trained for a higher number of epochs. This longer pre-training process allowed XLM-R models to gain a deeper understanding of linguistic patterns and representations, making them more effective in sentiment analysis for Sinhala. However, it is important to note that these monolingual models still demonstrate promising capabilities in capturing sentiment patterns in Sinhala text. The performance of these monolingual models can be further improved by pre-training the models on a larger Sinhala corpus for a higher number of epochs.

Figure 1 displays the row-wise normalized confusion matrix for sentiment analysis conducted using the XLM-R-large model with four classes. Based on the confusion matrix, we can deduce that the XLM-R-large model performs better in predicting the majority classes (Negative, Neutral, and Positive) than the Conflict class. There is a noticeable tendency for the model to misclassify instances labeled as Conflict as Negative at a relatively higher frequency. This misclassification pattern may be influenced by the class imbalance in the dataset, where the Negative class is the majority class with over 10,000 instances. The

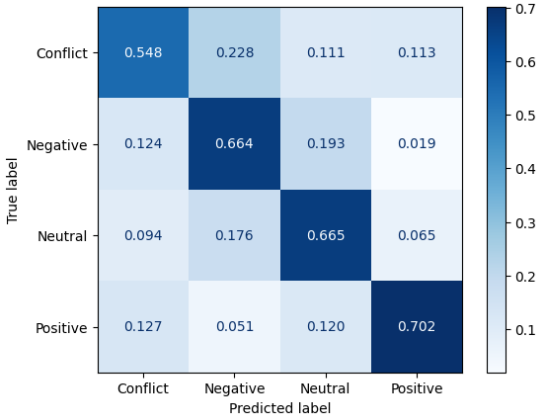

Figure 1: Normalized confusion matrix of XLM-R-large for sentiment analysis with four classes

model might have learned to favor the majority class, leading to more frequent misclassifications for the Conflict class. The class imbalance poses a challenge for the model to accurately distinguish between the classes, particularly affecting its ability to predict the minority class accurately.

6 Conclusion

This study evaluates the performance of various transformer models fine-tuned for sentiment analysis in the Sinhala language. This study marks the first experimentation of BERT and DistilBERT for sentiment analysis in Sinhala. The findings demonstrate that transformer models exhibit remarkable performance, even when fine-tuned using a small dataset. This outcome highlights the significant potential of transformer models in addressing challenges for languages with limited available resources. We also showed that the extensive pre-training process of the XLM-R models played a pivotal role in their superior performance compared to other models pre-trained for a single epoch.

In this study, we have made several contributions to the research community. We have made publicly available the pre-trained models of BERT, DistilBERT, and RoBERTa, along with the fine-tuned models of BERT, DistilBERT, RoBERTa, XLM-R-base, and XLM-R-large.

Additionally, three new datasets [4] have been released, which include a sentiment dataset comprising 5,037 news comments annotated with four classes, another dataset with 5,037 news comments annotated for three classes, and a large Sinhala news comments dataset containing 417,332 unannotated comments. These resources aim to foster further advancements and enable researchers to explore sentiment analysis in the Sinhala language more effectively.

These research outcomes contribute valuable insights to the field of sentiment analysis of low-resource languages and provide a foundation for future studies and applications. The utilization of transformer models, especially XLM-R-large, showcased promising results, indicating the potential for further advancements in sentiment analysis tasks for the Sinhala language.

7 Limitations

Despite the efforts to build and fine-tune transformer models for Sinhala sentiment analysis, several limitations remain.

Dataset Limitations: Although we used the OSCAR dataset for pre-training, the dataset size is limited to 2 GB, which may not fully encompass the diversity and complexity of the Sinhala language. This limited corpus may not provide sufficient exposure to a variety of linguistic expressions and dialects in Sinhala, thereby constraining the model's ability to generalize across different text types. Additionally, the fine-tuning dataset, consisting of comments from sources such as Lankadeepa online newspaper and GossipLanka news website, may introduce topic or sentiment biases that are not representative of broader Sinhala language use.

Annotation Limitations: Data annotation for sentiment analysis was conducted by two native Sinhala speakers, achieving an inter-annotator agreement score (Cohen's Kappa) of 0.794. While this indicates a good level of agreement, it also suggests some level of disagreement, which could lead to inconsistencies in sentiment labels and affect model performance. The annotations might contain subjective interpretations, especially in cases where sentiments are not explicit, and this could influence the accuracy and reliability of the final dataset.

Class Imbalance: Both datasets used in this study exhibit significant class imbalances, especially with the "Negative" class being dominant, which may bias the model towards negative predictions and reduce accuracy for underrepresented classes like "Conflict".

Limited Epochs for Pre-training: Given computational constraints, pre-training was limited to one epoch, which may not provide the models with enough iterations to fully capture the language patterns and features of Sinhala. However, the XLM-R model, which was already pre-trained for multiple epochs on a large corpus, produced better results due to its extensive pre-training process.

Limited Fine-Tuning: The models were fine-tuned using default configurations recommended in the original papers, without exploring alternative hyperparameters to identify the best setup. Although adjusting these settings could have enhanced model performance, this approach was not pursued due to limited computational resources.

These limitations indicate potential areas for future work, such as expanding the dataset, increasing annotation consistency, exploring additional model architectures, and conducting further experiments to enhance model generalizability for Sinhala language applications.

References

Julien Abadji, Pedro Ortiz Suarez, Laurent Romary, and Benoît Sagot. 2022. Towards a Cleaner Document-Oriented Multilingual Crawled Corpus.

P. D. T. Chathuranga, S. A. S. Lorensuhewa, and M. A. L. Kalyani. 2019. Sinhala Sentiment Analysis using Corpus based Sentiment Lexicon. In *2019 19th International Conference on Advances in ICT for Emerging Regions (ICTer)*, pages 1–7. IEEE.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised Cross-lingual Representation Learning at Scale.

Nisansa de Silva. 2019. Survey on Publicly Available Sinhala Natural Language Processing Tools and Research.

Piyumal Demotte, Lahiru Senevirathne, Binod Karunanayake, Udyogi Munasinghe, and Surangika Ranathunga. 2020. Sentiment Analysis of Sinhala

[4] https://github.com/bandaranayake/sinhala-sentiment-analysis

News Comments using Sentence-State LSTM Networks. In *2020 Moratuwa Engineering Research Conference (MERCon)*, pages 283–288. IEEE.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

Vinura Dhananjaya, Piyumal Demotte, Surangika Ranathunga, and Sanath Jayasena. 2022. BERTifying Sinhala - A Comprehensive Analysis of Pre-trained Language Models for Sinhala Text Classification. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 7377–7385, Marseille, France. European Language Resources Association.

C\'\icero dos Santos and Ma\'\ira Gatti. 2014. Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 69–78, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.

Yoav Goldberg and Omer Levy. 2014. word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method.

Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Stroudsburg, PA, USA. Association for Computational Linguistics.

Wenxiong Liao, Bi Zeng, Xiuwen Yin, and Pengfei Wei. 2021. An improved aspect-category sentiment analysis model for text sentiment analysis based on RoBERTa. *Applied Intelligence*, 51(6):3522–3533.

Weibo Liu, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, and Fuad E. Alsaadi. 2017. A survey of deep neural network architectures and their applications. *Neurocomputing*, 234:11–26.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach.

Nishantha Medagoda. 2016. Sentiment Analysis on Morphologically Rich Languages: An Artificial Neural Network (ANN) Approach. In pages 377–393.

Nishantha Priyanka Kumara Medagoda. 2017. *Framework for Sentiment Classification for Morphologically Rich Languages: A Case Study for Sinhala*. Ph.D. thesis, Auckland University of Technology.

Nishantha Medagoda, Subana Shanmuganathan, and Jacqueline Whalley. 2015. Sentiment lexicon construction using SentiWordNet 3.0. In *2015 11th International Conference on Natural Computation (ICNC)*, pages 802–807. IEEE.

Kostadin Mishev, Ana Gjorgjevikj, Irena Vodenska, Lubomir T. Chitkushev, and Dimitar Trajanov. 2020. Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers. *IEEE Access*, 8:131662–131682.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Stroudsburg, PA, USA. Association for Computational Linguistics.

Surangika Ranathunga and Isuru Udara Liyanage. 2021. Sentiment Analysis of Sinhala News Comments. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 20(4):1–23.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter.

Lahiru Senevirathne, Piyumal Demotte, Binod Karunanayake, Udyogi Munasinghe, and Surangika Ranathunga. 2020. Sentiment Analysis for Sinhala Language using Deep Learning Techniques.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, et al. 2019. HuggingFace's Transformers: State-of-the-art Natural Language Processing.

Guixian Xu, Yueting Meng, Xiaoyu Qiu, Ziheng Yu, and Xu Wu. 2019a. Sentiment Analysis of Comment Texts Based on BiLSTM. *IEEE Access*, 7:51522–51532.

Hu Xu, Bing Liu, Lei Shu, and Philip Yu. 2019b. BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis. In *Proceedings of the 2019 Conference of the North*, pages 2324–2335, Stroudsburg, PA, USA. Association for Computational Linguistics.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical Attention Networks for Document Classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1480–1489, Stroudsburg, PA, USA. Association for Computational Linguistics.