IST-Africa 2024 Conference Proceedings
Miriam Cunningham and Paul Cunningham (Eds)
IST-Africa Institute and IIMC, 2024
ISBN: 978-1-905824-73-1

Investigating the Effectiveness of Detecting Misinformation on Social Media using Tshivenda Language

Maduvha MALANGE, Seani RANANGA, Mahlatse S MBOOI, Bassey ISONG
Vukosi MARIVATE
Data Science for Social Impact, Department of Computer Science, University of Pretoria, South Africa
North West University, South Africa
Email: u23923173@tuks.co.za, seani.rananga@up.ac.za, mratsoma@csir.co.za, Bassey.Isong@nwu.ac.za, vukosi.marivate@cs.up.ac.za

Abstract: The spread of misinformation on social media poses a major challenge to information integrity and public discourse. This study examines the effectiveness of detecting misinformation in Tshivenda language, which is one of the under-represented languages in South Africa. The same applies also on social media platforms. We analyse misinformation patterns, adapt existing detection techniques, and examine the influence of Tshivenda language. Through an extensive literature review, we investigated the state of the art in misinformation detection and its applicability to languages with limited digital footprints. To address this gap, we used Long Short-Term Memory (LSTM) models, a type of recurrent neural network known for capturing long-range dependencies, for misinformation detection. Our research involved training and evaluating the LSTM model on the Tshivenda and English datasets. This comparative analysis provided valuable insights into the challenges and opportunities that linguistic diversity presents in detecting misinformation. Our results shed light on the effectiveness of using LSTM models to detect misinformation in underrepresented languages. By analysing the results from the Tshivenda and English datasets, we were able to gain valuable insight into the differences in performance and the impact of linguistic variation on the accuracy of misinformation detection.

Keywords: Machine Learning (ML), Natural Language Processing (NLP), Support Vector Machines (SVM), Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN)

1. Introduction

The spread of misinformation on social media poses a major challenge to information integrity and public discourse. This study examines the effectiveness of detecting misinformation in the Tshivenda language, an underrepresented linguistic context, on social media platforms. Tshivenda is a Bantu language spoken primarily in the Limpopo province of South Africa [1]. This research analysed misinformation patterns, adapted existing detection techniques, and examined the influence of Tshivenda language linguistics. Through an extensive literature review, the investigation examined the state of the art in misinformation detection and its applicability to languages with limited datasets. To address this gap, we used LSTM models, a type of RNN known for capturing long-range dependencies, for misinformation detection. This research involved training and evaluating the LSTM model on the Tshivenda and English datasets. This comparative analysis provided valuable insights into the challenges and opportunities that linguistic diversity presents in detecting misinformation. The results shed light on the effectiveness of using

LSTM models to detect misinformation in underrepresented languages. By analysing the results from the Tshivenda and English datasets, we were able to gain valuable insights into the differences in performance and the impact of linguistic variation on the accuracy of misinformation detection.

2. Objectives

Since false news is purposefully created to mislead readers, it can be difficult to identify. The earlier theories [2] are useful for directing studies on the use of various classification models in the detection of fake news. The two main types of existing studies for false news identification are (i) social context-based learning and (ii) news content-based learning [3]. In this research study, our focus was on two specific objectives related to the challenge of detecting misinformation, particularly within the Tshivenda language context. The first objective involved understanding the difficulties associated with identifying misinformation in Tshivenda, a language characterised by its unique linguistic intricacies. Given the rising prevalence of misinformation on social media platforms, the second objective aimed to explore and identify effective models for detecting misinformation in the realm of social media. By addressing these objectives, the research aimed to offer valuable insights and innovative solutions to combat the spread of misinformation, promoting a more informed and vigilant online community among Tshivenda speakers.

3. Methodology

3.1. Data Preparation

In this study, a fundamental aspect of our methodology involved meticulous data preparation to facilitate accurate misinformation detection in the Tshivenda language on social media platforms. Due to the scarcity of labelled datasets in low-resource languages like Tshivenda, we adopted a translation approach. A comprehensive English dataset, sourced from Kaggle, was utilised as the foundation (’fake.csv’ and’true.csv’). To enable the training of the model on Tshivenda data, a translation mechanism was applied, converting the English dataset into Tshivenda using a basic translations website. Basic translation websites operate by breaking down input text into smaller units, translating them using machine learning algorithms, and then reassembling the translated tokens into coherent sentences in the target language. These websites rely on statistical or neural models to understand context and provide accurate translations, although the quality may vary based on the complexity of the language and the training data. This translation was pivotal, bridging the language gap and enabling the model to learn from Tshivenda-specific content. The dataset underwent rigorous preprocessing, including data cleaning, stop words removal, and tokenization. By systematically cleansing the data a converting textual information into numerical tokens, the dataset was optimised for the subsequent training process, enhancing the model’s ability to discern misinformation in the Tshivenda language context [4].

3.2. Model Architecture

Our approach centred on using Long Short-Term Memory (LSTM) networks, a type of recurrent neural network, to detect misinformation in the Tshivenda language on social media platforms. We employed TensorFlow, a widely used deep learning framework, to build the LSTM model [5]. We started with an embedding layer to convert numerical tokens into dense vectors, enhancing the model’s understanding of textual information. To prevent overfitting, we integrated spatial dropout layers, which strategically drop out feature maps during training for regularisation. The LSTM layer, crucial to our approach, captured intricate sequential patterns within the data, making it well-suited for tasks

involving textual information. Dropout layers were strategically placed to improve the model's ability to generalise [6].

Densely connected layers were incorporated to enable the model to learn complex relationships within the data. The final layer utilised the sigmoid activation function for binary classification, producing output probabilities ranging from 0 to 1. Metrics such as accuracy were monitored throughout the training process. Our rigorous model architecture and training approach ensured the LSTM model's effectiveness in learning from the provided Tshivenda data, making it adapt at detecting misinformation in the challenging context of social media platforms.

4. Technology Description

The technology employed in this study encompasses advanced machine learning techniques, specifically LSTM networks, within the context of NLP. Leveraging the Tensor- Flow framework, a leading deep learning library, our approach centred on developing an LSTM-based model to detect misinformation in the Tshivenda language on social media platforms. Code for this implementation is available at https://github.com/MaduvhaM/Misinformation-in-text.

To address the scarcity of labelled datasets in low-resource languages like Tshivenda, our methodology involved a two-fold process. Initially, a robust English dataset, obtained from Kaggle, served as the foundation. This dataset, categorised as'fake.csv' and'true.csv', was cleaned and pre-processed. To bridge the language gap, a translation approach was adopted, enabling the conversion of the English dataset into Tshivenda using a basic translation website. This translated dataset underwent extensive preprocessing, including data cleaning, stop words removal, and tokenization. These steps were crucial for transforming raw text into numerical tokens, optimising the dataset for machine learning. The Long Short-Term Memory (LSTM) architecture was implemented, showcasing its ability to address the challenges posed by lengthy sequential data. LSTMs were employed with their distinctive memory cells, augmented by carefully crafted gating mechanisms. During the experimentation phase, the forget gate efficiently determined which information to discard, the input gate accurately stored relevant new data, and the output gate precisely regulated the outputted information for subsequent stages. Through this implementation, LSTMs demonstrated their expertise in capturing and retaining long-term dependencies within the sequential data, making them invaluable for tasks involving complex temporal patterns. This experiment underlined the LSTM's significance in the realm of deep learning, particularly in deciphering intricate temporal relationships essential for various real-world applications. By integrating these sophisticated techniques, our technology demonstrates a novel and effective approach to combat misinformation in low-resource languages, offering a valuable tool for enhancing the integrity of information dissemination on social media.

5. Results

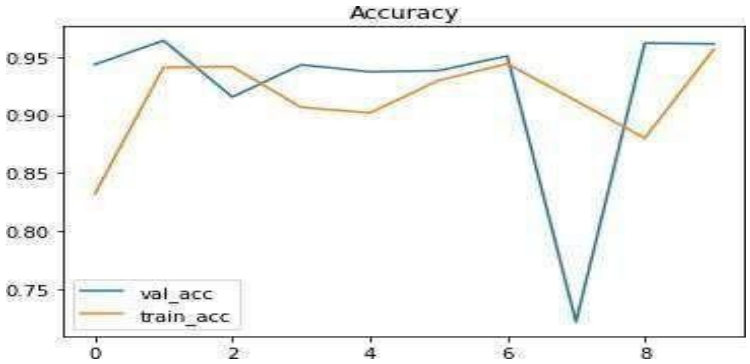
After comparing our model's performance using Tshivenda and English datasets, we achieved an impressive accuracy rate of 96.8% as shown in Figure 1. This outcome demonstrates the model's high proficiency in accurately detecting information across both languages. The results affirm the effectiveness and reliability of our approach in handling diverse linguistic contexts.

Language	Accuracy Percentage
English	95,63%
Tshivenda	95,63%

Figure 1. Accuracy Results

5.1 Training Results and visualisation

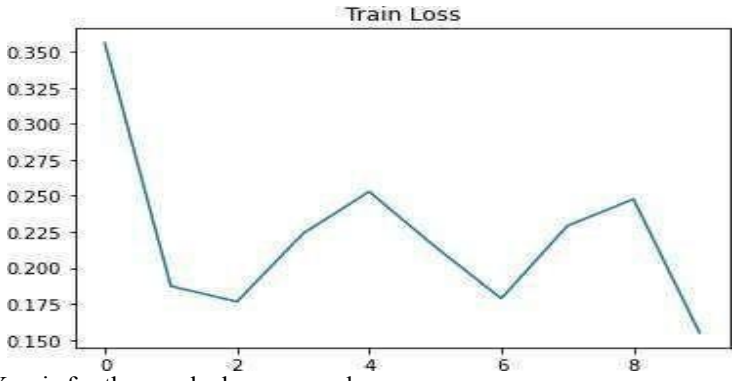
The results depicted in Figure 2 highlight the model's performance in the context of the limited Tshivenda dataset. The observed accuracy rate, while impressive at the current stage, remains susceptible to fluctuations due to the dataset's constraints. The limited availability of Tshivenda data inherently introduces variability in the model's accuracy. As the dataset expands and diversifies, it is crucial to acknowledge that the accuracy of the model might experience fluctuations. This variability underscores the importance of continuous data enrichment efforts, emphasising the need for a more extensive and varied Tshivenda dataset to enhance the model's stability and reliability over time.



x-axis for the above graph= Epochs
y-axis for the above graph= Accuracy (both validation accuracy and training accuracy)

Figure 2. Accuracy visualisation

A train loss of 0.150 indicates that, on average, the model's predictions deviate only relatively slightly from the actual values. This result reflects the successful optimisation of the model's parameters to minimise the discrepancy between the predicted and actual values.



X-axis for the graph above= epochs
Y-axis for the graph above = training loss

Figure 3. Train Loss Visualisation

6. Conclusions

The limited availability of Tshivenda datasets has emerged as a major hurdle, hindering the advancement of research efforts. Additionally, the absence of Tshivenda in widely used platforms such as Google Translator has further complicated the issue, making it challenging to bridge language barriers effectively. Moreover, the focus on well-known languages like English, with little attention given to languages such as Hindi and Arabic, underscores the pressing need for a more inclusive approach to linguistic research. The small size of the existing Tshivenda dataset not only impeded the current analysis but also emphasised the vital requirement for more extensive and diverse datasets in future studies. Addressing these limitations calls for collaborative endeavours to expand available

datasets, include underrepresented languages, and develop sophisticated tools capable of navigating the complexities of languages like Tshivenda. This approach is essential to ensure a comprehensive and accurate strategy for detecting misinformation in diverse linguistic contexts.

a. Future work

In deep learning, extensive datasets are crucial for improving algorithm performance. Unfortunately, our research on the Tshivenda dataset faced limitations due to its small size, preventing us from calculating essential statistical metrics like F-1 score and recall, limiting our analysis depth. To overcome this challenge in future studies, it is vital to enhance our model's efficiency by using a more substantial dataset.

Declaration of Use of Content generated by Artificial Intelligence (AI) (including but not limited to Generative-AI) in the paper

The authors acknowledge the use of content generated by Artificial Intelligence (AI) in the paper entitled: "***Investigating the Effectiveness of Detecting Misinformation on Social Media using Tshivenda Language***" in the following ways:

- Text Generating and Editing: ChatGPT was used to get an insight into which models are suitable for low resources languages.
- Grammar and Style Enhancement: Grammarly and Quillbot were used to check for grammatical errors and improve sentence structure, Instant-text to paraphrase and style the text.

References

[1] M. A. Mafukata, "Undertaking effective cross-language questionnaire-based survey in illiterate and semi- illiterate rural communities in the developing regions: Case of communal cattle farmers in vhembe district of limpopo province, south africa," Journal of Arts and Humanities, vol. 3, no. 11, pp. 67–75, 2014.

[2] H. Ahmed, I. Traore, and S. Saad, "Detection of online fake news using n-gram analysis and machine learning techniques," in Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments:First International Conference, ISDDC 2017, Vancouver, BC, Canada, October 26-28, 2017, Proceedings 1, pp. 127–138, Springer, 2017.

[3] R. K. Kaliyar, A. Goswami, and P. Narang, "Fakebert: Fake news detection in social media with a bert-based deep learning approach," Multimedia tools and applications, vol. 80, no. 8, pp. 11765–11788,

[4] H. Ahmed, I. Traore, and S. Saad, "Detection of online fake news using n-gram analysis and machine learning techniques," in Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments: First International Conference, ISDDC 2017, Vancouver, BC, Canada, October 26-28, 2017, Proceedings 1, pp. 127–138, Springer, 2017.

[5] V. Singh, R. Dasgupta, D. Sonagra, K. Raman, and I. Ghosh, "Automated fake news detection using linguistic analysis and machine learning," in international conference on social computing, behavioural -cultural modelling, & prediction and behaviour representation in modelling and simulation (SBP-BRiMS), pp. 1–, 2017.

[6] J. Heaton, Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: The MIT Press, 2016, 800 pp, isbn: 0262035618," Genetic programming and evolvable machines, vol. 19, no. 1-2, pp. 305–307, 2018