Machine Learning Research
2025, Vol. 10, No. 2, pp. 151–157
https://doi.org/10.11648/j.mlr.20251002.16

Research Article

Stance Classification in Afan Oromo Using Deep Learning Approaches

Dejene Teka*

Department of Information Technology, Mattu University, Oromia, Ethiopia

Abstract

Stance detection is an important task in natural language processing (NLP) that seeks to determine a speaker's or writer's position toward a given topic. While substantial progress has been achieved for major languages, low-resource languages such as Afan Oromo remain largely underexplored. This study introduces a deep learning–based approach for stance detection in Afan Oromo, leveraging a newly collected and annotated dataset of over one million sentences from social media platforms, particularly Facebook. Three deep learning models—Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Bidirectional LSTM (Bi-LSTM)—were implemented and evaluated. Among these, CNN achieved the highest accuracy of 85.9%, outperforming LSTM (81.4%) and Bi-LSTM (79.8%). The superior performance of CNN is attributed to its ability to capture local spatial features in text, which is particularly beneficial for short, informal social media posts. These results demonstrate the feasibility and effectiveness of deep learning techniques for stance detection in low-resource languages. Furthermore, the findings contribute to advancing language technologies for Afan Oromo and open pathways for future research in social media analysis, sentiment monitoring, and political discourse understanding in local contexts.

Keywords

Stance Detection, Afan Oromo, CNN, LSTM, Bi-LSTM, Deep Learning, Low-Resource Languages, Natural Language Processing

1. Introduction

1.1. Background of the Study

Afan Oromo is one of the prominent indigenous African languages, widely spoken across various regions of Ethiopia and extending into neighboring countries. It ranks as the second most widely spoken native language in sub-Saharan Africa. Within Ethiopia, it is the primary language of the Oromo people and is used by approximately 40% of the population. Additionally, it serves as a lingua franca in interactions with other ethnic groups. The language belongs to the Cushitic branch and is written using a Latin-based orthography. Its dialects vary significantly across geographical regions stretching from Tigray in the north to central Kenya in the south, and from Wallagga in the west to Harar in the east [1].

This study focuses on the challenges associated with Natural Language Processing (NLP) when interpreting words in context, particularly for languages like Afan Oromo. Many words have multiple meanings depending on how they are used, and some may share similar or entirely unrelated

*Corresponding author: dejutek@gmail.com (Dejene Teka), dejenetekahordofa@gmail.com (Dejene Teka)

Received: 6 August 2025; **Accepted:** 16 August 2025; **Published:** 19 December 2025

interpretations. Such ambiguity presents considerable obstacles for NLP systems, which must be capable of understanding the intended meaning in specific contexts. The need to contextualize word meanings accurately is critical to improving the performance of NLP tools [2].

Stance classification is a technique used to determine whether a speaker or writer is in favor of, against, or neutral toward a particular subject. It has practical applications in areas such as market analysis, political campaigns, and opinion mining on social media [3]. In the Ethiopian context, social media platforms such as Facebook and Twitter have become popular spaces for public discourse and political engagement. These platforms allow individuals to voice their opinions freely, contributing to the growth of politically active communities and digital advocacy groups [4].

The concept of stance relates to how individuals express their perspective or judgment regarding an issue. In the digital era, particularly on social media, stance classification has emerged as a valuable tool for assessing public opinion on socially and politically sensitive topics [5]. With millions of Ethiopian users on platforms like Facebook, such tools are increasingly relevant. Stance classification helps identify user positions on debates surrounding issues like climate change, gender equality, elections, and more.

Fundamentally, stance classification is a subfield of NLP focused on detecting attitudes or viewpoints shaped by cultural, social, and personal contexts. It is crucial in various applications, from automatically determining community opinions on political or religious matters to enhancing systems for recommendation, market trend prediction, and misinformation detection [6].

1.2. Statement of the Problem

With the growth of mobile technology and internet access, social media has become a major platform for public discourse in Ethiopia. While this presents opportunities for civic engagement and opinion sharing, it also creates challenges in understanding public attitudes toward policies and political developments. Stance classification—a subfield of natural language processing (NLP)—offers a method for analyzing users' support, opposition, or neutrality toward specific topics based on their textual expressions. [7, 8]

Although significant progress has been made in stance classification for resource-rich languages like English, such models are not readily transferable to under-resourced languages like Afan Oromo. [9, 10] This linguistic gap limits the ability to assess the perspectives of Oromo social media users on critical national issues such as political instability, displacement, unemployment, and rising living costs. Existing research largely excludes Afan Oromo, leaving a crucial gap in NLP-based stance analysis. This study seeks to address that gap by applying deep learning techniques to classify stances in Afan Oromo social media comments related to political parties and Oromo activists. [11, 12]

1.3. Research Questions

1) Which deep learning algorithms are most suitable for classifying Afan Oromo stances on social media?
2) Among the selected models, which one achieves the highest accuracy in classifying Afan Oromo stances?
3) How effectively do the developed models perform in the classification of Afan Oromo stances?

1.4. Objectives of the Study

General Objective

To design and develop a stance classification model for Afan Oromo social media texts using deep learning techniques.

Specific Objectives

1) To identify effective preprocessing techniques for Afan Oromo stance classification.
2) To build and evaluate different deep learning models for stance detection.
3) To assess the performance of local feature descriptors in the classification task.
4) To analyze and compare the performance of the developed models.
5) To draw conclusions and provide recommendations based on the findings.

2. Literature Review

2.1. Afan Oromo

Afan Oromo is a Cushitic language spoken by over 40 million people across Ethiopia and neighboring countries. It is written using the Latin-based Qubee script and has a rich oral and literary tradition. Despite historical marginalization, Afan Oromo is now an official language of Ethiopia and widely used in media and education [13, 14].

2.2. Natural Language Processing (NLP)

NLP is a field of computer science focused on enabling machines to understand and process human language. It supports applications such as sentiment analysis, machine translation, and voice assistants. With advances in computing and data availability, NLP has become central to extracting meaning from large volumes of text [15, 16].

2.3. Deep Learning

Deep learning uses neural networks to process large datasets and extract patterns for tasks like text classification, sentiment analysis, and language translation. It outperforms traditional methods by learning from raw data and has revolutionized fields such as computer vision and NLP [17, 18].

2.4. Stance Classification

Stance classification identifies a speaker's position (e.g., support, oppose, neutral) toward a target topic. It is widely used in politics, social media, and business. Techniques include machine learning, deep learning, and rule-based systems, with classification types ranging from binary to fine-grained categories [19].

2.5. Approaches to Stance Classification

Key approaches include feature-based machine learning, deep learning, and ensemble methods. Deep learning models like Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and especially Long Short-Term Memory (LSTM) networks are effective in capturing context and sequence dependencies in text. LSTM networks use memory cells to retain relevant information over time, making them well-suited for stance classification tasks [20, 21].

2.5.1. Deep Learning Approaches

Bidirectional LSTM (Bi-LSTM): Bi-LSTM processes input sequences in both forward and backward directions, enabling the model to learn from both past and future context. This dual processing improves the ability to capture long-term dependencies in text data. It typically uses features like word embeddings, character embeddings, and POS tags to enhance stance classification performance [22].

Convolutional Neural Networks (CNN): CNNs are widely used in NLP for feature extraction. They apply filters to capture patterns in input text and use pooling layers to reduce dimensionality. Although originally developed for image processing, CNNs are effective in text classification by identifying local dependencies and phrase structures in sentences [23].

2.5.2. Ensemble Learning Approaches

Ensemble learning combines multiple classifiers to improve prediction accuracy and model robustness. Methods like Random Forests and hybrid models (e.g., combining SVC, MLP, and Random Forest) have been successfully applied in stance detection on platforms such as Twitter, demonstrating improved classification results over single models.

2.5.3. Word Embedding

Word embedding represents words as dense vectors in a continuous space, capturing semantic relationships between them. Techniques like *Word2Vec* and *GloVe* are commonly used to generate embeddings by analyzing word co-occurrence patterns. Unlike one-hot encoding, word embeddings allow models to generalize better across similar contexts and deal with unseen words by mapping semantically related words to similar vectors [24].

3. Materials and Methods

3.1. Overview

This section outlines the methodology used in developing the Afan Oromo stance classification model. It details the system architecture, data collection, preprocessing, model building, and feature extraction processes.

3.2. System Architecture

The proposed model architecture consists of components for data collection, preprocessing, embedding generation, feature extraction, and classification, as illustrated in Figure 1.

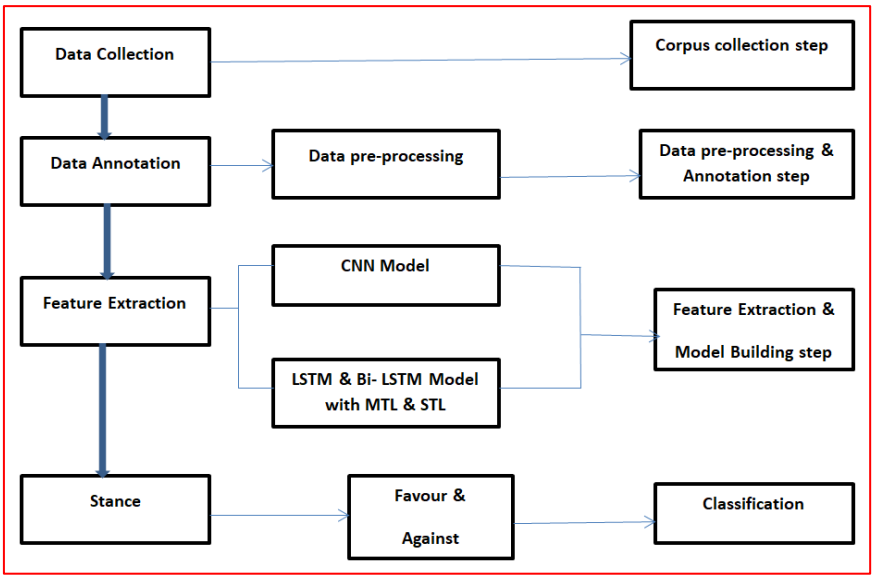


Figure 1. *System architecture.*

3.2.1. Data Collection

Around 1 million Afan Oromo sentences were collected through web scraping from platforms such as Facebook and YouTube. These texts were used to train a custom *Word2Vec* model, as no pre-trained embeddings for Afan Oromo were available. The model produced 100-dimensional vectors for approximately 100,000 unique tokens.

Data Annotation: Due to the lack of labeled Afan Oromo datasets, manual annotation was performed using simple Python tools in Jupyter Notebooks.

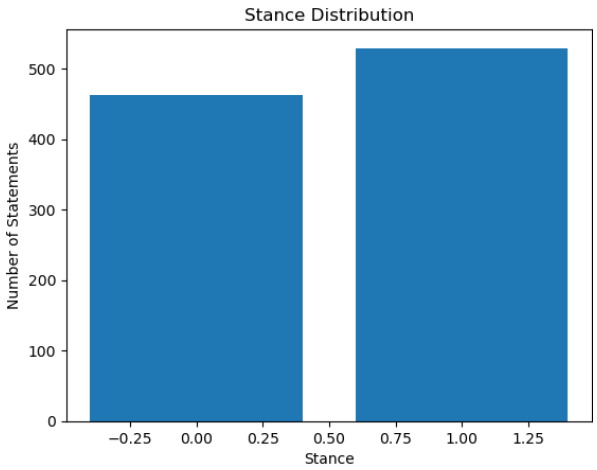
3.2.2. Dataset Preprocessing

1) *Normalization:* Standardizing text (e.g., lowercasing, removing punctuation).
2) *Tokenization:* Splitting sentences into individual words.
3) *Stop Words Removal:* Removing commonly used, low-value words like "kana," "isa," etc.

3.2.3. Feature Extraction and Model Building

The following steps were used:

1) *Embedding Layer:* A pre-trained Word2Vec model was used to initialize the embedding layer in Keras for better performance.
2) *Feature Extraction: A Convolutional Neural Network (CNN)* was used to extract features through convolution and pooling operations.
3) *Model Architecture:* The extracted features were passed to a *stacked Bidirectional LSTM (Bi-LSTM)* network to capture contextual dependencies in both directions.
4) *Model Training & Validation:* Standard machine learning practices including model selection, training, validation, and hyperparameter tuning were applied to optimize performance.



Out[13]:

	ID	text	Stance
0	1	koomiishiniin hoggansa ittisa ibiddaa fi balaa...	0
1	2	googil tajaajila of eeggannoo balaa lolaan dur...	0
2	3	hoggansi federaalaa fi naannoo amaaraa godina ...	0
3	4	balaan lolaa moozaambiik hamma yaadamee oli dh...	0
4	5	naannoleen sulula awaash lolaa mudatuun akka h...	0
...	...	...	...
987	988	itiyoophiyaan jioota 10niin darbanitti damee o...	1
988	989	godina shaawa bahaa aanaa gimbichuutti pirooje...	1
989	990	itiyoophiyaan jioota saglaan darbantii omisha ...	1
990	991	agarsisni huccu fi omisha huccu finfinne giddu...	1
991	992	sirna misooma lafa magaalaa fi bulchiinsa isaa...	1

Figure 2. *Analysis of annotated data.*

3.2.4. Proposed Model Training and Validation Loss

This section presents the training and validation loss history of the proposed model, offering insights into its learning behavior and generalization performance.

During training, the *training loss* reflects the model's performance on the training dataset, while the *validation loss* measures performance on a separate validation set that the model has not seen during training. These metrics are crucial for assessing both the model's ability to learn patterns from the training data and its capacity to generalize to new, unseen data.

To ensure effective learning, the model was optimized to *minimize the loss function*—typically *categorical cross-entropy* in classification tasks. However, it is equally important to monitor the validation loss to detect signs of *overfitting*, where the model performs well on training data but poorly on new input.

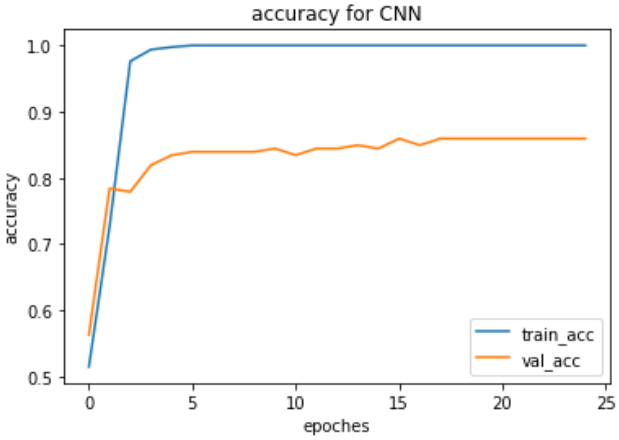


Figure 3. *Accuracy of CNN model.*

The *training and validation loss trends* were visualized using a line graph, where the *x-axis represents the number of epochs* and the *y-axis shows the loss value*. Ideally, both training and validation losses should decrease over time and eventually stabilize. A small gap between the two curves indicates good generalization, while a large gap suggests overfitting.

The graph in Figure 4 demonstrates the loss trajectories over the training epochs. As observed, the model gradually reduced its error on both datasets, indicating effective learning and convergence to a robust solution.

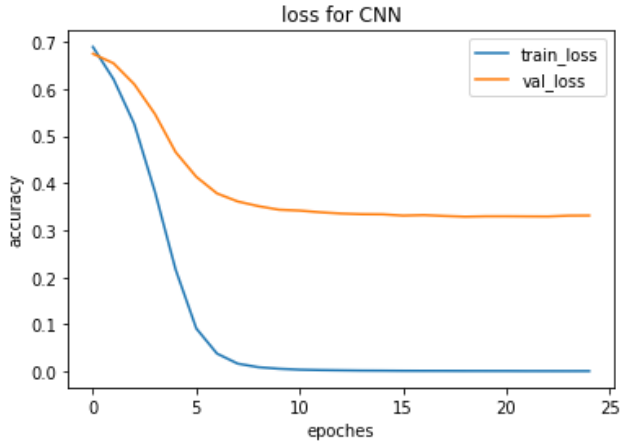


Figure 4. *Loss of CNN model.*

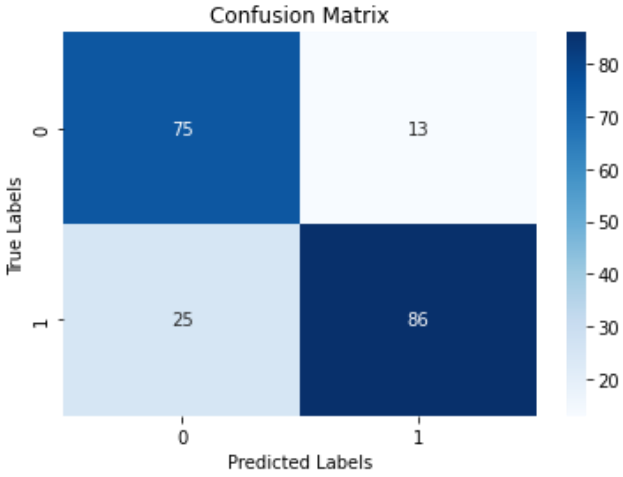


Figure 5. *Confusion Matrix.*

During the prediction phase, the trained language model takes a prompt or input text, which can be in the form of a comment, question, or statement. The model then processes this text and generates a response based on its training and understanding of language patterns, grammar, and context.

The model analyzes the input text and makes predictions about what the most likely next words or phrases should be, given the context of the prompt. It generates a response by selecting the most probable continuation of the text based on its training data.

It's important to note that the model's responses are not based on personal opinions or real-time information. The model generates responses based on patterns it has learned from the training data, which means it may not always provide accurate or up-to-date information.

In this research, the study explains the interpretation of predicted stance values in a sentiment classification model. A stance value of 0 indicates a favor stance, 1 indicates against stance. These values are assigned based on the model's analysis of comments.

4. Discussion

In the preceding section, we reported the classification performance of three deep learning models—*Convolutional Neural Network (CNN)*, *Long Short-Term Memory (LSTM)*, and *Bidirectional LSTM (Bi-LSTM)*—on the stance detection task using annotated Afan Oromo social media data. The models achieved classification accuracies of *85.9%* for CNN, *81.4%* for LSTM, and *79.8%* for Bi-LSTM.

These results demonstrate that the *CNN model outperformed both the LSTM and Bi-LSTM architectures*, achieving the highest classification accuracy. This suggests that CNN's *ability to extract local and spatial features* from text, particularly n-gram patterns through convolutional filters, made it more suitable for capturing discriminative stance-related features in this dataset. Unlike RNN-based models that rely on sequential processing, CNNs are known for *efficiently identifying short-range dependencies*, which can be especially effective for stance detection tasks where key stance indicators often appear in localized patterns (e.g., hashtags, strong sentiment words).

The LSTM and Bi-LSTM models, while still achieving competitive results, were slightly less accurate. One possible explanation is that *LSTM-based models are designed to capture long-range dependencies* in text, which might not have been critical for this specific dataset. Moreover, the increased complexity and number of parameters in Bi-LSTM may have led to *overfitting*, particularly if the training data did not contain enough examples to justify such complexity.

It is important to emphasize that while accuracy provides a general overview of performance, *additional evaluation metrics*—such as *precision*, *recall*, *F1-score*, and *ROC-AUC*—are essential for a more comprehensive assessment, especially in tasks where *class imbalance* may exist. For instance, high accuracy may mask poor performance on minority classes. A future extension of this work should incorporate these metrics for a more balanced evaluation.

Furthermore, several *limitations* need to be acknowledged. First, although CNNs performed well, their fixed-size input and lack of sequence modeling could limit their applicability to more complex text tasks. Second, the dataset used was

collected from a single domain (social media posts), which may not generalize well to other domains such as news articles or academic writing. Lastly, annotation inconsistencies or noise in user-generated content may also affect model performance.

Future research directions include exploring *transformer-based models* (e.g., BERT, RoBERTa) which have demonstrated superior performance across many NLP tasks, including stance detection. In addition, *data augmentation*, *multi-task learning*, or the incorporation of *external linguistic resources* (such as sentiment lexicons or dependency parsing) may further enhance the model's performance and robustness.

Table 1. *Comparison of CNN, LSTM, BI-LSTM model.*

Model	Accuracy (%)
CNN	85.9
LSTM	81.4
Bi-LSTM	79.8

5. Conclusion and Recommendation

5.1. Conclusion

This study aimed to develop and evaluate deep learning models for stance detection in the Afan Oromo language, utilizing a custom-annotated dataset collected from social media sources. The models compared included Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Bidirectional LSTM (Bi-LSTM).

Among the models, *CNN outperformed the others with an accuracy of 85.9%*, followed by LSTM (81.4%) and Bi-LSTM (79.8%). These results indicate that CNN's strength in extracting local spatial features made it more effective for this specific task, likely due to the nature of short, informal, and often repetitive patterns in social media text.

This research highlights the potential of applying deep learning techniques to low-resource languages like Afan Oromo, where stance detection can help in understanding public opinion, monitoring political sentiment, and detecting misinformation in online platforms. Furthermore, the annotated dataset and training methods used here provide a foundation for future research in similar underrepresented languages.

5.2. Recommendation

1) *Model Enhancement:* While CNN performed well, combining CNN with attention mechanisms or transformer-based models like BERT may yield even better results. Future work should explore hybrid models or pre-trained multilingual transformers fine-tuned on Afan Oromo.
2) *Dataset Expansion:* Increasing the size and diversity of the dataset—especially including data from other platforms (e.g., Twitter, Telegram)—would enhance the generalization capabilities of the models.
3) *Language Resources Development:* There's a critical need for developing more linguistic tools and resources for Afan Oromo (e.g., tokenizers, stemmers, sentiment lexicons), which would greatly benefit various NLP tasks beyond stance detection.
4) *Real-world Application:* These models can be deployed in systems for media monitoring, hate speech detection, or early warning systems for conflict-prone discourse. Partnerships with local media, civil society, and tech organizations would help turn this research into practical tools.
5) *Interdisciplinary Collaboration:* Future studies should involve collaboration between linguists, data scientists, and domain experts to ensure cultural relevance, ethical data use, and effective deployment of NLP solutions in local contexts.

Abbreviations

NLP	Natural Language Processing
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
Bi-LSTM	Bidirectional Long Short-Term Memory
BERT	Bidirectional Encoder Representations from Transformers

Author Contributions

Dejene Teka is the sole author. The author read and approved the final manuscript.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

References

[1] Wondimu. Tegegne, The development of written Afan Oromo and the appropriateness of Qubee, Latin Script, for Afan Oromo writing. Historical Research Letter, p. 28, 2016.

[2] Abdo Ababor. Abafogi, "Enhanced Word Sense Disambiguation Algorithm for Afaan Oromoo.," Information Engineering and Electronic Business, pp. 41-50, February 8, 2023. https://doi.org/10.5815/ijieeb.2023.01.04

[3] Saif M., Parinaz Sobhani, and Svetlana Kiritchenko. Mohammad, "Stance and sentiment in tweets," Transactions on Internet Technology (TOIT), pp. 1-23, 2017. https://doi.org/10.48550/arXiv.1605.01655

[4] A. Tazeb, "The Framing of Political Uprising on Social Media: The Case of," Journalism and Communication, pp. 1-120, 2017.

[5] D. Küçük and F. C. SIGIR, "Stance detection: Concepts, approaches, resources," Proceedings of the 44th International ACM, 2021.

[6] "Stance detection D. Küçük and F. C. SIGIR, "Concepts, approaches, resources," Proceedings of the 44th International ACM;, 2021.

[7] S. O. Edosomwan, "The history of social media and its impact on business," Article in The Journal of Applied Management & Entrepreneurship, vol. 16, 2011.

[8] A. Arora, P. Nakov and I. Augenstein M. Hardalov, "ASurvey on Stance Detection for Mis- and Disinformation Identification," arXiv preprint, pp. 1-19, 8 May 2022.

[9] D. Dou and D. Lowd J. Ebrahimi, "A joint sentiment-target-stance model for stance classification in tweets," Proceedings of COLING 2016, the 26th, 2016.

[10] H. Luqman and M. Ahmed N. Alturayeif, "A systematic review of machine learning techniques for stance detection and its applications," Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016.

[11] T. Tuunanen, C. E. Gengler, M. Rossi, W. Hui, V. Virtanen and J. K. Peffers, "Design Science Research Process: A Model for Producing and," International Conference on Design Science Research in Information Systems, https://doi.org/10.2753/MIS0742-1222240302

[12] Kareem, et al. Darwish, "Predicting online islamophobic behavior after# ParisAttacks.," The Journal of Web Science, vol. 3, pp. 34-52, 2018.

[13] D. Küçük and F. Can, "Stance detection: A survey," Journal of ACM Computing, vol. 53, pp. 1-37, 2021.

[14] Galataa. Abbabaa, "QAACCESSA QABIYYEE FAARUU QAALLUU MACCAA: AANAA SIBUU SIREETTI. Diss. Yuunivarsiitii Haramaayaa," 2018.

[15] Tom, et al. Young, "Recent trends in deep learning based natural language processing," IEEE, pp. 55-77, 2018.

[16] Nader, et al. Tavaf, "GRAPPA-GANs for parallel MRI reconstruction." arXiv preprint arXiv: 2101.03135 (2021).".

[17] Maryem, et al Rhanoui, "A CNN-BiLSTM model for document-level sentiment analysis," *Machine Learning and Knowledge Extraction*, pp. 832-847., 2019.

[18] Yann, Yoshua Bengio, and Geoffrey Hinton LeCun, "Deep learning," pp. 436-444, 2015.

[19] S. M., & Bravo-Marquez, F. Mohammad, "Stance and sentiment in tweets. ACM Transactions on Internet Technology (TOIT), 17(3), 26.," 2017.

[20] A., & Paul, S. Bhatia, "Stance classification in online debate portal," *In 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pp. 1-5, 2016.

[21] C., Huang, L., Qiu, X., & Hu, X. Sun, "LSTM-based deep learning models for non-factoid answer selection. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics," *Long Papers*, pp. 1142-1151.

[22] Y., Du, N., Yuan, M., & Huang, X. Li, "A Survey on Bi-Directional LSTM: Architecture, Modeling, and Applications," *Journal of Ocean University of China*, pp. 513-524., 2019.

[23] D., Wang, X., & Fang, Y. Zhang, "A comprehensive survey on convolutional neural networks." *Journal of Machine Learning Research*, pp. 3199-3230., 2017.

[24] T., Chen, K., Corrado, G., & Dean, J. Mikolov, "Efficient estimation of word representations in vector space," 2013.