International Journal of Advanced Engineering and Technology
www.allengineeringjournal.com
ISSN: 2456-7655
Received: 28-05-2023, Accepted: 13-06-2023, Published: 28-06-2023
Volume 7, Issue 1, 2023, Page No. 23-29

Development and evaluation of a framework for detecting hate speech and abusive language in Zambia using machine learning

Clement Mulenga Sinyangwe[1], Douglas Kunda[2], William Abwino Phiri[3]
[1] Department of ICT, Chalimbana University, Lusaka, Zambia
[2] Vice Chancellor, ZCAS University, Lusaka, Zambia
[3] Vice Chancellor, Chalimbana University, Lusaka, Zambia

Abstract
The advent of artificial intelligence (AI) has revolutionized various fields, including information technology, intelligent transportation systems, virtual personal assistants, robotic surgery, and natural language processing (NLP) applications. However, along with the numerous benefits brought about by technological advancements, there are also drawbacks, such as the widespread dissemination of abusive language, fake news, and hate speech, which can easily be propagated in the digital world. Social media platforms like Facebook and Twitter have played a significant role in the rapid spread of rumors, conspiracy theories, hatred, xenophobia, racism, and prejudice. The misuse of technology has not only influenced public opinion but also impacted religious views worldwide, enabling targeting of individuals based on various attributes. Zambia's social media landscape has witnessed a dynamic shift, particularly following the transition of the government in 2021, which has led to greater freedom of expression but also an upsurge in hate speech and abusive language associated with political, ethnic, and religious divisions. The freedom of expression (FoE) in Zambia has facilitated the sharing of diverse views and ideas, contributing to development, democracy, and dialogue. However, this freedom has also led to the proliferation of hate speech on various online platforms, including social media. Despite the efforts of governments, the technology industry, and individual researchers to address the issue of hate speech, challenges persist. Legislative measures have been attempted to suppress hate speech, but their effectiveness is often limited. The main objective of this study was to develop and evaluate the framework for detecting hate speech and abusive language in Zambia. Cross-Industry Standard Procedure for Data Mining (CRISP-DM) methodology, a commonly used method for overseeing data science projects, was used to perform this study. precision, recall, and F1 score was used to evaluate the framework. Gradient-boosted decision tree was picked over the other algorithms (KNeighbors Classifier, logistic Regression, Random Forest, Decision Tree and Naïve Bayes) because apart from being a powerful machine learning algorithm that has become increasingly popular in recent years, especially in tasks such as classification and regression.

Keywords: Development, evaluation, framework, detecting, hate speech, abusive language, machine learning

Introduction

Artificial intelligence (AI) has revolutionized various fields, but it also has drawbacks like the widespread dissemination of abusive language, fake news, and hate speech. Social media platforms like Facebook and Twitter have played a significant role in the rapid spread of rumors, conspiracy theories, hatred, xenophobia, racism, and prejudice. The misuse of technology has influenced public opinion and religious views worldwide, enabling targeting of individuals based on various attributes. Zambia's social media landscape has experienced a dynamic shift, particularly following the government's transition in 2021. (Abdullah *et al*., 2020) [4] stated that COVID-19 pandemic has further exacerbated the negative impact of abusive language, fake news, and hate speech. The World Health Organization (WHO) has highlighted the threat posed by misinformation and hate speech in the context of the pandemic. As more people join online social networking platforms, such as Twitter, it becomes easier for negative comments and hateful messages to proliferate. To address these challenges, there is a pressing need to develop an effective framework for detecting and combating hate speech and abusive language in Zambia's online environment. This research aims to develop and evaluate a machine learning-based framework that can automatically identify and mitigate instances of hate speech and abusive language, leveraging the unique linguistic and cultural nuances of the Zambian context to enhance accuracy, efficiency, and scalability in hate speech detection (Salim., 2020) [3]. By successfully developing this framework, we can contribute to a safer and more inclusive online environment in Zambia, empowering individuals, governments, and online platforms to take proactive measures against the spread of hate speech and abusive language (Lata., 2021) [8].

Problem Statement

The freedom of expression (FoE) in Zambia has enabled the sharing of diverse views and ideas, but it has also led to the proliferation of hate speech on various online platforms, including social media (Siame, 2019) [13]. According to Chen & Delany (2017), hate speech can cause physical harm and spread discrimination, making it difficult to maintain a safe and inclusive online environment. Automatic detection of hate speech is crucial, but the problem is complex due to the wide range of hate speech types and the lack of robustly annotated hate speech corpora. Despite efforts by governments, the technology

industry, and researchers, challenges persist in suppressing hate speech.

The Zambia Information and Communications Technology Authority (ZICTA) reports an exponential increase in cyber offenses and complaints (ZICTA, 2020) [19]. The urgent need is to develop an effective framework for detecting and mitigating hate speech and abusive language in Zambia's online environment. This framework should use machine learning techniques to accurately identify and classify hate speech instances, considering the diverse nature of hate speech and existing detection approaches. By addressing this problem, we can contribute to a safer and more inclusive digital space in Zambia, safeguarding individuals from the harmful consequences of hate speech and promoting a culture of respectful online communication. However, Zambia, we have 72 languages of which some of them are closely related and use some common phrases and words that have a completely different meaning. E.g., one would use a Mambwe word "inyele" meaning human hair. The same word translated in Nyanja would mean an abusive term describing someone very horny. This may affect the accuracy of the results especially the local one.

Problem Analysis

language dictionary on detection model accuracy

The use of a language dictionary in machine learning can significantly impact the accuracy of a detection model. It provides a comprehensive list of words and phrases relevant to the problem being addressed, but it can also limit the flexibility of the detection model, making it less effective at identifying novel or emerging forms of problematic language. In Zambia, the linguistic diversity of the country means that data sets for specific languages may be limited or non-existent. By using language dictionaries, it is possible to expand the data sets used to train detection models, improving their accuracy for specific languages.

Effects of language domain on detection model accuracy

The domain of the language used in machine learning can significantly affect the accuracy of a detection model. The language used in different domains, such as social media, news articles, scientific papers, or legal documents, can vary significantly in terms of vocabulary, grammar, and syntax. If the detection model is trained on a dataset that is representative of the language used in a particular domain, it is likely to perform well in that domain. However, if the detection model is not representative of the language used in the target domain, its accuracy may be lower.

Limited Dataset on detection model accuracy

The size and quality of the dataset used to train a detection model can have a significant impact on its accuracy. If the dataset is limited in size or quality, it can have a negative impact on its accuracy. Researchers can collect more data from a diverse range of sources, such as social media, news articles, and spoken conversations, to improve the accuracy of detection models by providing a wider range of language patterns and reducing bias towards specific languages or domains. Additionally, a limited dataset may make it difficult to detect rare or emerging forms of problematic language that are not well-represented in the dataset.

Deep learning classifiers used

1. **Gradient-boosted decision trees:** Gradient-boosted decision trees are an ensemble learning method that combines multiple decision trees to create a powerful predictive model. They have been applied in various domains, including sentiment analysis, document classification, logistic regression, random forest, and text classification tasks. Gradient-boosted decision trees have been used in studies to classify movie reviews based on sentiment, outperforming traditional machine learning algorithms (Si., 2017).

2. **K-Nearest Neighbours (KNN):** K-Nearest Neighbors (KNN) classifiers classify new instances based on the majority vote of neighboring data points, and KNN has been widely used for document classification tasks. (Neloy., 2019).

3. **Logistic regression models:** Logistic regression models the relationship between input features and the probability of a certain outcome, achieving good accuracy in identifying spam emails (Campbell *et al*., 2019) [28].

4. **Random Forest:** Random Forest is an ensemble learning method that combines multiple decision trees to make predictions, and it has been successfully used in image classification (Xu., 2012).

5. **Naive Bayes:** Naive Bayes is a probabilistic classification algorithm based on Bayes' theorem and the assumption of feature independence. It calculates the probability of a data point belonging to a certain class based on the probability distributions of its features. Naïve Bayes classifiers have shown competitive performance in accurately identifying spam emails (Leung., 2007) [15].

Objective

The main objective of this study was to develop and evaluate the framework for detecting hate speech and abusive language in Zambia

Litereture Review

Framework for hate speech detection using Machine Learning Approach

Palimote., (2021) [29] developed a machine learning framework for detecting hate speech using a Twitter dataset containing 24,784 tweets. The dataset was reduced to two columns using feature extraction, with tweet columns containing hate speech and class columns containing 0,1 and 2, respectively. The model was trained using support vector machine and random forest classifier, achieving 95% and 99% accuracy. The model was deployed to the web using Python Flask for easy evaluation and testing. The experimental results showed that the proposed system outperformed other methods in classifying text as hate speech.

Deep Learning Framework for Automatic Detection of Hate Speech

Duwairi, and Quwaider (2021) [27] studied a deep learning framework for automatic detection of hate speech embedded in Arabic tweets. They trained and tested deep networks on

the ArHS dataset, which contains 9833 tweets. The researchers also investigated performance on two existing Arabic hate speech datasets, resulting in a combined dataset of 23,678 tweets. The study found that CNN outperformed other models in binary classification, with 81% accuracy. In ternary classification, both CNN and BiLSTM-CNN models achieved the best accuracy of 74%. In multi-class classification, CNN-LSTM and BiLSTM-CNN models achieved the best results with 73% accuracy. In the combined dataset, the CNN-LSTM and BiLSTM-CNN models achieved the best accuracy of 65% in binary classification, 67% in ternary classification, and 65% in multi-class classification (Simonyan., 2015) [2]

Framework for hate speech detection using deep convolutional neural network

Pradeep *et al*., (2020) [12, 26] studied a framework for hate speech detection using deep convolutional neural networks. They found that rapid internet user growth led to cyber issues like cyberbullying and hate speech. Hate speech focuses on protected aspects like gender, religion, race, and disability, and can lead to unwanted crimes. To monitor hate speech on Twitter, an automated system was developed using Deep Convolutional Neural Network (DCNN). The proposed model captures tweet text with GloVe embedding vector, achieving precision, recall, and F1-score values of 0.97, 0.88, and 0.92, outperforming existing models.

Framework for automated hate speech detection and span extraction

Zhou *et al*., (2022) [11, 25] conducted a study on automated hate speech detection and span extraction in underground hacking and extremist forums. The researchers used a dataset of posts from Hack Forums and Stormfront and Incels.co, as well as a Twitter hate speech dataset, to train a multi-platform classifier. The study found that a classifier trained on multiple sources of data does not always improve performance compared to a mono-platform classifier. The researchers fine-tuned BERT and used span prediction and sequence labeling approaches to extract hateful spans, achieving an F1-score of at least 69%.

Framework detection of fake news and hate speech

Wubetu and Ayodeji (2022) [10, 24] conducted a study on detecting fake news and hate speech in Ethiopian languages. The study analyzed the optimal approaches and their relationship with dataset type, size, and accuracy. Deep learning (DL) approaches were recommended for Ethiopian languages to improve performance across various social media platforms. Combining DL and machine learning (ML) approaches with a balanced dataset can improve detection and combating performance (Goodfellow *et al* 2016) [1]. Combating hate speech and fake news is a pressing societal issue, and automatic fact or claim verification has gained interest. However, their results remain unsatisfactory, requiring further research in this area. Fake news and hate speech messages spread negativity about sex, caste, religion, politics, race, disability, and sexual orientation, making it difficult to detect and combat.

Framework for hate speech detection using deep convolutional neural network

A study by Roy *et al*., (2020) [23], developed a framework for identifying hate speech using deep convolutional neural networks. The study found that the rapid increase in internet users led to issues like cyberbullying and hate speech. Hate speech targets protected characteristics like gender, religion, race, and disability, and can result in unintentional crimes. To monitor hate speech, an automated system was created using Deep Convolutional Neural Network (DCNN) and achieved the best precision, recall, and F1-score values of 0.97, 0.88, and 0.92 using tweet text and GloVe embedding vector.

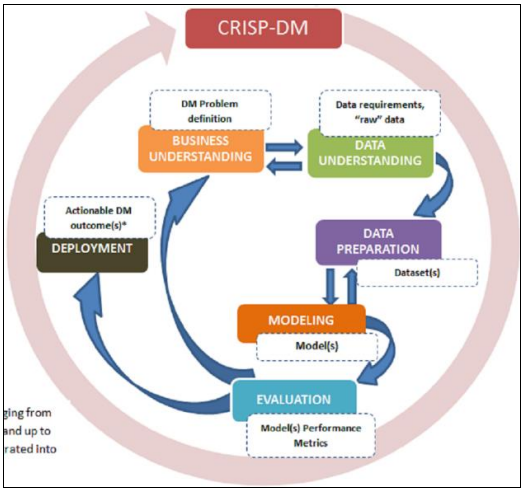
Framework for detection of hate speech in videos using machine learning

Unnathi (2019) [9, 22] studied the detection of hate speech in videos using machine learning. The project categorized videos into normal or hateful categories using a crawler and a Speech-to-Text converter. Four models were trained using three different feature sets from the dataset. The Random Forest Classifier model performed best in categorizing films. The increase in hate speech has led to disputes and cyberbullying. Advancements in machine learning and deep learning have prompted researchers to investigate and implement solutions to hate speech (Yoon *et al*., 2020) [5] Machine learning approaches are now applied to textual data, making a method for identifying hate speech in videos essential due to the widespread use of video sharing websites.

Framework for detection of hateful comments on social media

Essa (2022) [20] conducted research on identifying offensive comments on social media, proposing a unique method called the Naive Bayes classifier. The Naive Bayes method had an accuracy of 62.75%, but the Neural algorithm improved it to 87%. The study found that social media usage has increased, but some users abuse it by inciting hatred, contradicting its goal of fostering close relationships. A comprehensive scientific strategy for identifying, measuring, and classifying hateful remarks on social media is currently lacking. The majority of cases are unreported due to social factors like victimization fear and psychological effects of hateful remarks. This lack of clarity hinders efforts to create processes and regulations to reduce the negative impacts of hate speech on social media, ultimately making platforms less useful as communication tools.

Methodology



CRISP-DM methodology stages
Connolly & Begg, 2014 [7]

The Cross-Industry Standard Procedure for Data Mining (CRISP-DM) methodology, a commonly used method for overseeing data science projects, was used to perform this study. It can be modified for deep learning projects even if it was originally created for data mining tasks (Connolly & Begg, 2014).

This methodology consists of six phases of the CRISP-DM methodology applicable to a deep learning project. Below is a summary of the six stages;

Table 1

Stage	Activity
Business Understanding	The project goals and requirements are defined. This phase is critical for deep learning projects, as it helps to ensure that the model being built will meet the business needs. For example, a deep learning project aimed at predicting customer churn could begin by clearly defining what constitutes churn and why it is important to the business.
Data Understanding	Data is gathered and analyzed to determine its quality, quantity, and suitability for the project. This is especially important for deep learning projects, as the effectiveness of a model depends on the quality and quantity of dataset being used for training. Data preparation and cleaning are typically required in this phase.
Data Preparation	This phase involves converting data into suitable formats for training a deep learning model. This may include tasks such as data normalization, feature engineering, and data augmentation.
Modeling	In this phase, a machine learning model is built and trained using the prepared data. The effectiveness of the model is evaluated on a validation set to determine if further tuning is required.
Evaluation	In this phase, the quality of the model is assessed on a test set to determine how well it generalizes to new data. This phase is critical for deep learning projects, as overfitting can be a common issue
Deployment	In this phase, the quality of the model is assessed on a test set to determine how well it generalizes to new data. This phase is critical for deep learning projects, as overfitting can be a common issue.

Tthe CRISP-DM methodology provides a structured approach to deep learning projects, ensuring that the project is well-defined, the data is of high quality, and the model is deployed effectively.

Findings and Discussions

Evaluating the performance of a hate speech and abusive language detection framework is crucial to ensure accurate identification of such content. Metrics like precision, recall, and F1 score can be used to assess the framework's performance. Precision measures the proportion of true positive detections among all detected instances, while recall measures the proportion of true positive detections among all actual positive instances in the dataset. The F1 score, the harmonic mean of precision and recall, is a common performance metric in binary classification problems, including Gradient Boosting. It measures the model's ability to correctly classify positive cases and identify all positive cases. In Gradient Boosting, the F1 score evaluates the model's ability to correctly classify positive and negative cases based on a chosen threshold value, which is the probability value above which an observation belongs to the positive class.

To calculate the F1 score, firstly the computation of the precision and recall of the model using the following formulas:

Precision = True Positives / (True Positives + False Positives)

Recall = True Positives / (True Positives + False Negatives)

True Positives (TP) refer to the number of positive cases that the model correctly classified, False Positives (FP) refer to the number of negative cases that the model incorrectly classified as positive, and False Negatives (FN) refer to the number of positive cases that the model incorrectly classified as negative.

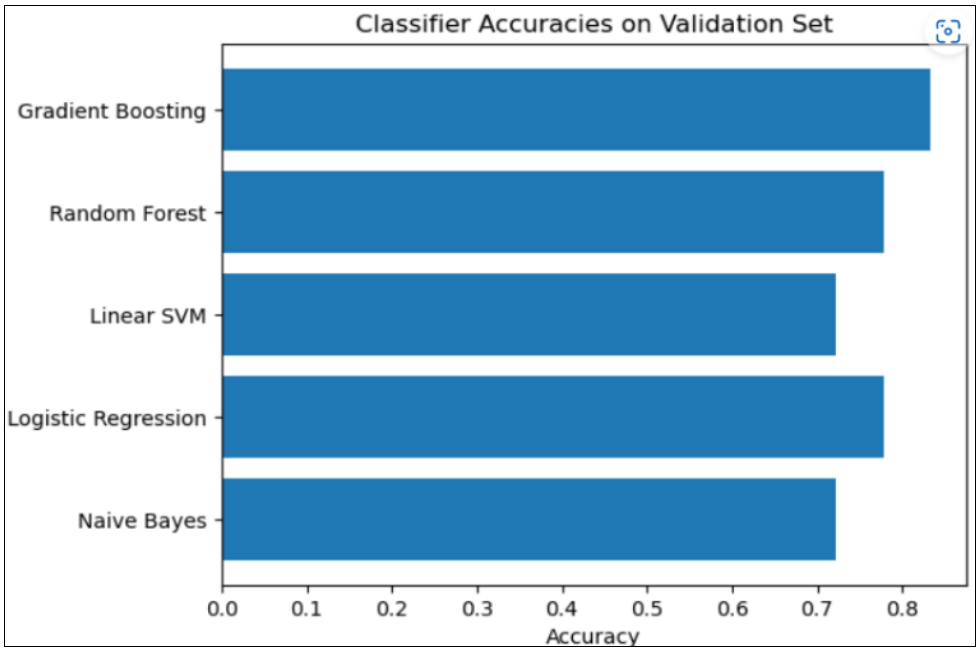
Below is the summary of the evaluation of the performance for the five algorithms used in training the model

```
import matplotlib.pyplot as plt

classifiers = [MultinomialNB(), LogisticRegression(), LinearSVC(), RandomForestClassifier(), GradientBoostingClassifier()]
classifier_names = ['Naive Bayes', 'Logistic Regression', 'Linear SVM', 'Random Forest', 'Gradient Boosting']
accuracies = []

for clf in classifiers:
    clf.fit(X_train, y_train.ravel())
    y_pred = clf.predict(X_vad)
    accuracy = accuracy_score(y_vad, y_pred)
    accuracies.append(accuracy)

fig, ax = plt.subplots()
ax.barh(classifier_names, accuracies)
ax.set_xlabel('Accuracy')
ax.set_title('Classifier Accuracies on Validation Set')
plt.show()
```




```
Naive Bayes
Accuracy: 0.7222222222222222
Logistic Regression
Accuracy: 0.7777777777777778
Linear SVM
Accuracy: 0.7222222222222222
Random Forest
Accuracy: 0.8333333333333334
Gradient Boosting
Accuracy: 0.8333333333333334
```

Upon obtaining the computations of the precision and recall values, the F1 score was calculated as follows:
F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

```
Gradient Boosting
Accuracy: 0.8333333333333334
Classification Report:
              precision    recall  f1-score   support

     class 0       1.00      0.50      0.67         4
     class 1       0.81      1.00      0.90        13
     class 2       0.00      0.00      0.00         1

    accuracy                           0.83        18
   macro avg       0.60      0.50      0.52        18
weighted avg       0.81      0.83      0.80        18
```

The F1 score ranges from 0 to 1, with 1 indicating perfect precision and recall, and 0 indicating poor performance. In logistic regression, the F1 score along with other performance metrics like accuracy, area under the ROC curve, and confusion matrix was then used to evaluate the model's performance and in selecting the best threshold value for classification.

```
# making the confusion matrix
# A confusion matrix is a technique for summarizing the performance of a classification algorithm.

def set_confusion_matrix(clf, X, y, title):
    plot_confusion_matrix(clf, X, y)
    plt.title(title)
    plt.show()
```

Gradient-boosted decision tree (GBDT) is a powerful machine learning algorithm that is popular for classification and regression tasks. It can handle both numerical and categorical features, making it more versatile in handling complex datasets like local languages. GBDT also handles outliers and missing values more effectively than other algorithms and learns complex decision boundaries by combining multiple decision trees. Popular implementations like XGBoost, LightGBM, and CatBoost have won machine learning competitions and become the go-to algorithm for many data scientists. GBDT can help organizations and individuals extract insights from their data, make informed decisions, and improve performance and outcomes in their respective domains.

Empirical findings from studies have shown that Gradient Boosting is the top-performing algorithm for detecting hate speech and offensive language in social media. However, Varsha *et al*. (2020) [20] achieved high accuracy and F1 scores, but their algorithm differed from ours. These findings emphasize the importance of evaluating machine learning frameworks for detecting hate speech and offensive language using various metrics and highlight the challenges researchers face, such as language variations and the need for domain-specific datasets and techniques.

Recommendations and Conclusion

This study concentrated on establishing whether hate speech and abusive language exists in Zambia and as such, the focus was building a detection model based on the local language dataset. However, during the process, some challenges of local language dictionaries and language domain emerged. Even though, these challenges did not hinder the functioning of the model, it however affected only the accuracy of the results. This implied that the study was therefore not entirely exhaustive in detecting a lot of offenses on a good number of online social media platforms especially those made in typical local languages. Therefore, the following recommendations were made for future consideration;

Creating a complete local data dictionary

Addressing Zambia's linguistic diversity is crucial for improving hate speech and abusive language detection algorithms. Future research should focus on creating a comprehensive data dictionary for Zambian languages, which are spoken by over 72 tribes and 102 dialects. This would provide a stable foundation for training and validation, enabling detection algorithms to capture nuances and context-specific elements of hate speech and abusive language across various Zambian languages. A comprehensive data dictionary would enable academics and programmers to continuously update and improve models, supporting the development of a more precise and effective system for addressing hate speech and abusive language, promoting a secure online environment for all users.

A local language dictionary is essential for preprocessing data for deep learning models, using techniques like word embeddings, language modeling, and named entity recognition. Transfer learning, a machine learning technique, uses knowledge learned from pre-trained models to improve performance on different tasks, saving time and resources. Reinforcement learning (RL) is a subfield of machine learning that focuses on software agents learning to take actions in an environment to maximize a cumulative reward signal. Despite its success, RL can be challenging due to sample inefficiency, exploration-exploitation tradeoffs, and reward shaping.

Conclusion

The study aimed to address ethical concerns in Zambian online media content and develop a machine learning-based framework for detecting hate speech and abusive language. It conducted a comprehensive analysis, created a multilingual dataset, and developed a machine learning-based framework. The framework considers the linguistic nuances of Zambia's local languages, enhancing its effectiveness in identifying and flagging problematic content. The evaluation results showed the framework's effectiveness and robustness in detecting hate speech and abusive language on Zambian platforms. To improve accuracy, models should be updated and trained on new data regularly. To ensure fairness and accuracy, varied and representative data should be included in the training process. Machine learning-based hate speech and abusive language detection algorithms can be effective tools for combating hate speech and fostering online inclusivity.

References

1. Goodfellow I, Bengio Y, Courville A. Deep learning. MIT press, 2016.
2. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint, 2015, 14.
3. Salim CER, Suhartono D. A systematic literature review of different machine learning methods on hate speech detection. JOIV: International Journal on Informatics Visualization, 2020.
4. Abdullah A, Alqurashi E, Alanazi M, Alaskar A, Alabdulkarim S. The effect of e-learning on academic performance among university students during the COVID-19 pandemic in Saudi Arabia: A mediating analysis. Education and Information Technologies, 2020.

5. Yoon J, Jordon J, van der Schaar M, Hu X. Deep learning in healthcare: Recent advances and challenges. IEEE Journal of Biomedical and Health Informatics, 2020.
6. Malmasi S, Dras M. Deep learning for natural language processing: An overview of recent developments. Annual Review of Linguistics,2020:6:435-455. doi: 10.1146/annurev-linguistics.
7. Connolly TM, Begg CE. Database Systems: A Practical Approach to Design, Implementation, and Management 6th edition. Pearson Education Limited, 2014.
8. Lata S. Hate speech detection framework from social media content in Ethiopia. International Journal of Computer Applications, 2021.
9. Unnathi H. Framework for detection of hate speech in videos using machine learning. In 2019 International Conference on Communication and Signal Processing (ICCSP), 2019.
10. Wubetu A, Ayodeji J. A framework for detection of fake news and hate speech using deep learning. In Proceedings of the 13th International Conference on Computer and Automation Engineering, 2022.
11. Zhou Y, Pete S, Hutchings G. A framework for automated hate speech detection and span extraction. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), 2022.
12. Pradeep S, Asis K, Tapan KS, Xiao Y. A framework for hate speech detection using deep convolutional neural network. In Proceedings of the 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020.
13. Siame KK. Social media and religious tolerance in Zambia. Journal of African Media Studies, 2019.
14. Chen CC, Delany SJ. Predicting sentiment polarity in reviews using multi-level text analysis. In Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining, 2012.
15. Leung KM. Naive bayesian classifier. Polytechnic University Department of Computer Science/Finance and Risk Engineering,2007, 123-156.
16. Di Cicco M, Potena C, Grisetti G, Pretto A. Automatic model-based dataset generation for fast and accurate crop and weeds detection. In 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2017, 5188-5195. IEEE.
17. Bhowmik NR, Arifuzzaman M, Mondal MRH. Sentiment analysis on Bangla text using extended lexicon dictionary and deep learning algorithms. Array,2022:13:100123.
18. Liu X. Evaluation of the Accuracy of Artificial Intelligence Translation Based on Deep Learning. Mobile Information Systems, 2022.
19. ZICTA. ZICTA Annual Report Annual Report, Advancing the Nation to a Digital Society, 2020, 2–2.
20. Varsha P, Manish J, Prasad AJ, Monica M, Tanmay J. Using Machine Learning for Detection of Hate Speech and Offensive Code-Mixed Social Media text. North Maharashtra University, Jalgaon MS. India, 2020.
21. Essa. Detection of Hateful Comments on social media. Rochester Institute of Technology, 2022.
22. Unnathi. Detection of Hate Speech in Videos Using Machine Learning. San Jose State University. USA., 2019.
23. Roy PK, Tripathy AK, Das TK, Gao XZ. A framework for hate speech detection using deep convolutional neural network. IEEE Access,2020:8:204951-204962.
24. Wubetu, Ayodeji. Detection of fake news and hate speech for Ethiopian languages: a systematic review of the approaches. Demilie and Salau Journal of Big Data,2022:9:66 https://doi.org/10.1186/s40537-022-00619-x
25. Zhou Pete, Hutchings. Automated hate speech detection and span extraction in underground hacking and extremist forums University of Cambridge, Cambridge CB2 1TN, UK, 2022.
26. Pradeep, Asis, Tapan, Xiao. A Framework for Hate Speech Detection Using Deep Convolutional Neural Network. Vellore Institute of Technology, Vellore 632014, India, 2020.
27. Duwairi, Quwaider. A Deep Learning Framework for Automatic Detection of Hate Speech Embedded in Arabic Tweets. King Fahd University of Petroleum & Minerals, 2021.
28. Campbell BC, Majoie CB, Albers GW, Menon BK, Yassi N, Sharma G, *et al*. Penumbral imaging and functional outcome in patients with anterior circulation ischaemic stroke treated with endovascular thrombectomy versus medical therapy: a meta-analysis of individual patient-level data. The Lancet Neurology,2019:18(1):46-55.
29. Palimote, Gaage. An Improve Framework for hate speech detection using Machine Learning Approach. Department of Computer Science, Kenule Beeson Saro wiwa Polytechnic, Bori, River State, Nigeria, 2021.