©NIJCRHSS Publications

Open Access
Print ISSN:2992-6106
Online ISSN:2992-5789

NEWPORT INTERNATIONAL JOURNAL OF CURRENT RESEARCH IN HUMANITIES AND SOCIAL SCIENCES (NIJCRHSS)

Volume 4 Issue 2 2024

https://doi.org/10.59298/NIJCRHSS/2024/4.2.1823

Analyzing the Application of Machine Learning in Detecting Hate Speech: A Review

Ezeaku Florence Uzoaji

Department of Business Management Kampala International University Uganda

ABSTRACT

Social media platforms offer avenues for fostering anonymous online connections, discussions on diverse topics like culture, politics, and community life. However, the proliferation of hate speech poses a pressing challenge for society, individuals, policymakers, and researchers alike, both on the continent and globally. Addressing this issue necessitates comprehensive studies to identify and combat hate speech effectively. This paper conducts a systematic review of literature in this domain, concentrating on methodologies such as word embedding, machine learning, deep learning, and cutting-edge technologies. Specifically focusing on the past six years of research, this review highlights gaps, challenges, and advancements in hate speech detection techniques. Additionally, it delves into limitations, algorithmic selection dilemmas, data collection complexities, cleaning challenges, and outlines future research pathways in this critical area.

Keywords: Hate Speech Detection, Machine Learning, Social Media Platforms, Text Analysis, Algorithm Selection.

INTRODUCTION

The rapid adoption of social media has revolutionized communication, leading to extensive information dissemination. In Ethiopia, social media serves as a primary information source, but its unverified content poses challenges [1]. Notably, during the Tigray conflict, social media played a role in civil disruptions, road closures, displacements, and loss of lives [2]. Instances of inciting violence against ethnic minorities circulated online [3]. The conflict also led to media shutdowns and arrests of journalists [4]. Globally, hate speech and fake news have raised concerns, prompting some countries to balance free speech with preventing its dangerous evolution [5]. UNESCO outlines strategies like media ethics education, conflict-sensitive reporting, social media oversight, reporting mechanisms, and fighting impunity for hate crimes [6]. This scientific investigation explores hate speech's contemporary landscape and Ethiopian language-specific challenges. Addressing hate speech is critical globally, impacting lives and historical narratives [7]. Public awareness and concerted efforts are essential for citizens' safety [8]. Ethiopia faced turmoil due to misinformation on social media, affecting education and trade [9]. A scientific solution derived from comprehensive research is crucial. This survey identifies current challenges, integrates existing knowledge, and guides future research.

Hate Speech

Hate speech encompasses any communication—verbal, written, or through actions—that employs offensive or discriminatory language targeting individuals or groups based on attributes like religion, ethnicity, race, gender, or other defining factors [10, 11]. It disrupts community harmony and perpetuates discrimination. In response to rising incidents of hate speech, the UN introduced the UN Strategy and Plan of Action on Hate Speech in 2019, aiming to address root causes and societal impacts through thirteen commitments, including victim support, social media awareness, and education [12]. Social media's pivotal role in daily communication, information dissemination, and entertainment amplifies hate speech's impact. Studies exist to mitigate this issue, and major platforms like Facebook, Twitter, and YouTube employ models to distinguish hate speech from genuine content [13]. However, diverse languages and interpretations pose ongoing challenges in hate speech identification, which often targets multiple aspects of identity, including gender, religion, race, and disability [14].

Open Access
Print ISSN:2992-6106
Online ISSN:2992-5789

Gender-based Hate Speech

Expressions disseminating, endorsing, or provoking hatred based on gender fall under gender-based hate speech, primarily targeting women and girls [15]. This form, known as sexist hate speech, condemns and undermines women, fostering fear and mistrust toward them in society. Technological advancements and social media accessibility have escalated violence against women and girls [16]. Online violence against women, primarily facilitated by social networks, significantly impacts their personal lives and professional careers [16]. Research indicates that this harassment could drive women to join extremist groups [17]. Additionally, [18] identified high instances of cyber harassment against women on social media platforms. This research paper delves into the pervasive issue of online bullying, particularly affecting young adults who experience severe forms of harassment. Recognizing the urgency, the paper aims to fill research gaps and propose solutions to combat this escalating problem.

Religious Hate Speech

Hatred directed at religions such as Islam, Hinduism, and Christianity constitutes a severe form of hate speech, impacting groups rather than individuals [19]. Extremist individuals within these religions face negative stereotypes, discrimination, and online abuse, particularly the rising trend of anti-Muslim abuse [20]. The internet's role as an amplifier intensifies existing discourses, fostering polarization [20]. Social media platforms serve as a breeding ground for illegal activities, fueling misunderstandings, intolerance, and heated religious debates, leading to tensions among followers [21]. This phenomenon spans Europe, Asia, and Africa [21]. Hate speech online exacerbates religious tensions, fueling extreme animosity [22]. History has shown the dangers when religion is manipulated for political motives, especially in Ethiopia's diverse religious landscape, where conflicts can have extensive consequences. Rigorous scientific research is imperative to prevent such incidents from recurring.

Racist Hate Speech

Racist hate speech targets a person or group based on appearance and fosters the belief in the inferiority of certain racial groups [23]. Such expressions often occur at the international level, influenced by governmental attitudes and leadership, varying in frequency and impact. The concerning rise of individuals propagating racial hatred on social media signifies a growing population that rejects equality [24]. Identifying these individuals requires a scientific approach to address moral degradation and promote equality. This compilation of studies offers crucial insights to tackle this urgent issue.

Hate speech on disability

Disability hate speech seeks to dehumanize individuals with physical or mental disabilities, viewing disability as a societal construct akin to race or gender [25]. Disabilities can arise from various conditions, including medical errors, accidents, or natural causes. Unfortunately, online social media users perpetuate hate speech that harms individuals with disabilities, impacting their living conditions and inclusion within communities [26]. This hinders their interaction and creates discord. Recognizing equality as a fundamental human value is vital, necessitating educational efforts to promote self-improvement, unity, and the disregard of disabilities.

Stages of Hate Speech

According to [27], hate speech follows a four-stage process. Initially, the "influence stage" emerges, characterized by heightened social media activity after an event, intensifying the spread of hate speech. This is succeeded by the "intervention stage," where the event's impact diminishes over several days. Subsequently, a gradual decline leads to the "response stage," where the impact reaches zero. The final "rebirth stage," represented by a dashed line in the illustration, is an optional phase where hate speech may resurface or not based on subsequent occurrences' nature and impact [27].

Hate Speech Techniques

The detection of hate speech has been predominantly approached through three primary methodologies: Keyword-Based Techniques, Machine Learning Techniques, and Deep Learning Techniques, often used individually or in hybrid forms. Researchers have extensively investigated the prevalence and usage patterns of these methodologies in hate speech detection. The overall composition is given below:

Keyword-Based Technique

As discussed in [28], the keyword-based method is a foundational approach in identifying hate speech. This technique involves scanning text for potentially offensive keywords through a predefined dictionary, often collected from diverse social media platforms like Facebook, Twitter, blogs, forums, and YouTube. However, this approach has limitations. While these keywords are associated with repugnant actions, their context and meaning can evolve over time, rendering them insufficient for comprehensive hate speech detection. Content devoid of explicit slurs might not be flagged, posing a challenge to this method. For instance, the phrase "umu anumanu e bilie" might have a benign literal meaning but could hold different implications in various contexts, evolving over time based

Open Access
Print ISSN:2992-6106
Online ISSN:2992-5789

on interpretation. Additionally, as highlighted in [29], keyword-based techniques struggle to detect hate speech conveyed through metaphors or slang lacking explicit hate keywords. For instance, the slang expression "umu inya amaghi egwuregwu," literally translating to "donkeys are not careful when playing," may hold a different, potentially contentious meaning in a political or religious context.

Machine Learning Technique

Described in [30], machine learning is a scientific methodology harnessed by computer systems that leverage algorithms and statistical models to efficiently perform tasks. Unlike explicit programming, it relies on patterns and data to operate. As a subset of artificial intelligence, machine learning algorithms develop a mathematical model based on training data, allowing them to make predictions or decisions without specific task-oriented programming. The primary goal is to create a classifier or regression model by learning from a training dataset, followed by evaluating its performance using test data. Machine learning encompasses supervised, unsupervised, and semi-supervised learning methodologies.

Deep Learning Technique

Described in [31], deep learning is a machine learning technique that emulates human learning by using examples. This approach employs neural networks to tackle complex problems. Through deep learning, computer models can directly perform classification tasks from inputs such as images, text, or sound, achieving notably high levels of accuracy. Training these models requires extensive labeled data and intricate, multi-layered neural network structures [32].

Hybrid Technique

This method is used to overcome the limitations inherent in a single approach. By merging two or more approaches into a hybrid method, leveraging their strengths to complement each other, it presents a promising solution.

Keyword-Based Technique

The technique involves the collection and categorization of keywords within a specific context, assessing their frequency within a document to understand its content. Illustrated below: In their study as detailed in [33], an analysis was conducted on the Ethiopic Twitter Dataset for Abusive Speech in Amharic. The study aimed to gather data for training linguistic models in language identification tasks and to examine the prevalence of specific keywords linked to abusive language. The textual data encompassed the Amharic, Tigrinya, and Ge'ez languages, totaling approximately three million tweets from 154,477 users between mid-August 2014 and 2019. Native speakers compiled 99 hate speech and 48 offensive speech keywords for the Amharic language, sourced from Facebook, Twitter, and YouTube comments. Excluding data from 2015 due to encoding issues, the assembled 147 keywords were categorized into hate speech and offensive speech. The analysis revealed an increasing trend in both the number of Amharic tweets and the prevalence of tweets containing offensive keywords over time.

Machine Learning Techniques

The study outlined in [34] centered on assessing data quality, focusing on hate speech within online discourse. It commenced by utilizing a hate speech lexicon from Hatebase.org, compiling words and phrases identified as hate speech by Internet users. Employing the Twitter API, the researchers searched for tweets containing lexicon terms, amassing a sample from 33,458 Twitter users, resulting in about 84.4 million tweets. Of these, 25,000 tweets containing lexicon terms were manually labeled by Crowd Flower workers. Workers were tasked with categorizing each tweet as hate speech, offensive yet not hateful, or neither offensive nor hate speech, considering the context in which the words were used. Using an inter-coder agreement approach involving three or more workers, 24,802 samples were considered, as most tweets lacked consensus among coders. From these, 5% were categorized as hate speech by majority consensus, with 1.3% deemed as such without opposition from coders. As the study employed stringent criteria for hate speech classification, most tweets were labeled as offensive language, with a minority classified as non-offensive. The study employed the Porter stemming algorithm and TF-IDF for analysis, evaluating tweet quality through modified Flesch-Kincaid Grade Level and Flesch Reading Ease scores. Five classical algorithms—logistic regression, naive Bayes, decision trees, random forests, and linear SVMs—were employed, with Grid Search used to optimize parameters. To prevent overfitting, 5-fold cross-validation was employed, highlighting logistic regression and linear SVM as the top-performing models. Ultimately, logistic regression was chosen, exhibiting an overall precision of 0.91, recall of 0.90, and an F1 score of 0.90. This study's focus on data quality aimed to comprehend the context in which offensive language and hate speech impact race, religion, and societal identity.

Deep Learning Techniques

The work presented in [35] aimed to create an automated Amharic Hate Speech Posts and Comments Detection Model using Recurrent Neural Networks (RNNs). It began with a literature review covering various approaches to online hate speech, including legal perspectives addressing hate speech on social media across different continents and countries. A dataset of 30,000 manually collected posts and comments from prominent activists and news pages was annotated into hate and free speech categories. The data underwent pre-processing through cleaning and normalization techniques. An RNN was developed utilizing LSTM and Gated Recurrent Unit (GRU), leveraging word n-grams for feature extraction and word2vec for word representation. The LSTM and GRU models were trained and tested on the dataset, split into training, validation, and test sets in an 80:10:10 ratio. Experimentation with different parameters on the GRU and LSTM-based RNN models using word2vec feature representation yielded a superior test accuracy of 97.9% achieved by RNN-LSTM. In [36], a study on Bangla hates speech detection on social media employed an attention-based recurrent neural network. Gathering 7,425 comments from Facebook, the data was divided into 80% for training and 20% for testing. Encoder–decoder-based machine learning models, including attention mechanism, LSTM, and GRU-based decoders, were utilized to predict hate speech categories. Among these algorithms, the attention-based decoder attained the highest accuracy, reaching 77%. These studies utilized different recurrent neural network architectures for hate speech detection in Amharic and Bangla languages, achieving high accuracies through their respective approaches.

Hybrid Techniques

The study detailed in [37] focused on detecting and classifying Afaan Oromo hate speech on social media. Collecting 12,812 data instances from Afaan Oromo language-based Facebook accounts, various machine learning algorithms were applied, including classical (SVM and NB), ensemble (RF and XGBoost), and deep learning (CNN and BiLSTM), utilizing different feature extraction techniques such as BOW, TF-IDF, Word2vec, and embedding layers. Two experiments were conducted for eight and two-class classifications. SVM with Word2vec achieved 82% accuracy for eight-class classification among classical and ensemble machine learning algorithms, while BiLSTM with pretrained Word2vec obtained 84% accuracy for the same classification. Additionally, BiLSTM with pretrained Word2vec attained a performance result of 0.88% accuracy. Moreover, the authors suggested further research directions in hate speech detection for audio, video, emoji, and memes, particularly in multilingual contexts. In the same vein, the study analyzed hate speech identification in Hinglish language using transformer models mBERT and IndicBERT for feature selection, focusing on code-mixed tweets. They tested traditional machine learning classifiers (SVM, LR, RF, NB, KNN) on translated and transliterated Devanagari script using mBERT embeddings. Subsequently, they experimented with the Deep Neural Network (DNN) model. Their results showed that their model outperformed existing methods, achieving a 73% accuracy for hate speech identification in Hinglish. The mBERT model and traditional machine learning classifiers exhibited better performance compared to IndicBERT for hate content detection in Hindi and English datasets. Both studies demonstrated the effectiveness of various machine learning techniques, including hybrid approaches, in detecting hate speech across different languages and social media contexts [37].

Comparative Analysis

The paper employs diverse methods and languages to detect hate speech content. It primarily utilizes classical machine learning algorithms and integrates them with deep learning methods either individually or in conjunction with other machine learning techniques. The amalgamation of deep learning with other algorithms enables the creation of models' adept at effectively identifying hate speech. Despite the substantial data needs of deep learning algorithms, their showcased accuracy in the reviewed studies establishes them as a favoured option for hate speech detection.

Findings

Each language harbors its unique nuances, making a machine learning algorithm effective in one language potentially ineffective in another. However, by meticulously observing linguistic intricacies and conducting thorough research, gathering and organizing data for such studies is achievable. This survey primarily delves into reviewing research related to the identification and anticipation of hate speech in the Amharic language. It accentuates the imperative for more extensive research on Ethiopian languages. The survey illuminates significant gaps within the field, notably the absence of benchmarks and various natural language processing tools tailored to these languages. Furthermore, it highlights a deficiency in standardized data annotation practices, often leading to research outputs that may appear hurried and lacking depth. Despite the volume of studies conducted, a palpable dearth persists in the body of knowledge concerning Amharic and other Ethiopian languages.

CONCLUSION

This paper conducts a comprehensive survey focusing on the automated detection of hate speech. It explores diverse aspects such as terminology, hate speech stages, detection methods, features like word2vec and embeddings,

Open Access
Print ISSN:2992-6106
Online ISSN:2992-5789

and model performance. The study scrutinizes various research methodologies, emphasizing their data preprocessing techniques and highlighting encountered challenges. Notably, the surveyed studies often neglect figurative language and visual content, despite their significant role in conveying hate messages. An important challenge identified pertains to data collection, primarily aggregating and evaluating data at the individual level, potentially overlooking hate speech manifested through bullying or offenses against ethnic minorities. To address these issues, the paper recommends the creation of a benchmark dataset for hate speech identification, aiming to facilitate better comparison across different features and methodologies.

REFERENCES

1. Assefa, M. (2020). Role of social media in Ethiopia's recent political transition. Journal of Media and Communication Studies, 12, 13–22.
2. Kingawa, Edemealem & Tasew, Kafte & Hailu, Seble & Sholaye, Muluken & Girmaw, Mekashaw & Teklemarkos, Senait & Feyisa, Terefe & Bitew, Abiyot. (2023). HATE SPEECH DETECTION USING MACHINE LEARNING: A SURVEY. Vol. 17. 88-109.
3. Shifa, M., & Leibbrandt, M. (2022). Spatial Inequality in Sub-Saharan Africa. Forthcoming in *African Geographical Review.*
4. Fred Harter. (2022). Ethiopia Gets Tough on Journalists Since Tigray Conflict. https://www.voanews.com/a/ethiopiagets-tough-on-journalists-since-tigrayconflict-/6683980.html
5. Brüggemann, S., Robert Prosser, A., & Ru, S. (2022). A scientific basis for a policy fighting fake news and hate speech. https://doi.org/10.24989//ocg.v.342
6. Kingawa, Edemealem & Tasew, Kafte & Hailu, Seble & Sholaye, Muluken & Girmaw, Mekashaw & Teklemarkos, Senait & Feyisa, Terefe & Bitew, Abiyot. (2023). HATE SPEECH DETECTION USING MACHINE LEARNING: A SURVEY. Vol. 17. 88-109.
7. Michael A. Peters (2022) Limiting the capacity for hate: Hate speech, hate groups and the philosophy of hate, Educational Philosophy and Theory, 54:14, 2325-2330, DOI: 10.1080/00131857.2020.1802818
8. Getahun, Surafel, Success and Failure of National Dialogue Selected Countries Cases Study: General Lesson to Ethiopia (June 8, 2023). Available at SSRN: https://ssrn.com/abstract=4473420 or http://dx.doi.org/10.2139/ssrn.4473420
9. freedomhouse. (2021). infrastructural limitations restrict access to the internet or the speed and quality of internet connections. https://freedomhouse.org/country/ethio pia/freedom-net/2021
10. Lumen Learning (2023). Introduction to Psychology. Lumen Learning https://courses.lumenlearning.com/waymaker-psychology/.
11. Nazmine, & Manan, Khan & Tareen, Hannan Khan & Noreen, Sidra & Tariq, Muhammad. (2021). Hate Speech and social media: A Systematic Review. Turkish Online Journal of Qualitative Inquiry. 12. 5285-5294.
12. UN. (2019). United Nations Strategy and Plan of Action on Hate Speech. United Nations Report, May, 1–5.
13. Burnap, P., & Williams, M. (2015). Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modeling for Policy and Decision Making: Machine Classification of Cyber Hate Speech. Policy & Internet, 7. https://doi.org/10.1002/poi3.85
14. Seglow, Jonathan. (2016). Hate Speech, Dignity and Self-Respect. Ethical Theory and Moral Practice. 19. 10.1007/s10677-016-9744-3.
15. Sękowska-Kozłowska, Katarzyna & Baranowska, Grażyna & Gliszczyńska-Grabias, Aleksandra. (2022). Sexist Hate Speech and the International Human Rights Law: Towards Legal Recognition of the Phenomenon by the United Nations and the Council of Europe. International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique. 35. 10.1007/s11196-022-09884-8.
16. Violence, O. G. (2023). GENDER-BASED And Its Impact On The Civic Freedoms of Women GENDER-BASED And Its Impact On The Civic Freedoms of Women (Issue March). https://www.icnl.org/wpcontent/uploads/Online-Gender-BasedViolence-report-final.pdf
17. Edwards, S. S. M. (2017). Cyber-Grooming Young Women for Terrorist Activity: Dominant and Subjugated Explanatory Narratives BT - Cybercrime, Organized Crime, and Societal Responses: International Approaches (E. C. Viano (ed.); pp. 23–46). Springer International Publishing. https://doi.org/10.1007/978- 3-319-44501-4_2
18. Rahman, G., et al. (2018) Spatial and Temporal Variation of Rainfall and Drought in Khyber Pakhtunkhwa Province of Pakistan during 1971-2015. Arabian Journal of Geosciences, 11, Article No. 46.https://doi.org/10.1007/s12517-018-3396-7

19. Kiper, T. (2013) Role of Ecotourism in Sustainable Development. In: Advances in Landscape Architecture, IntechOpen, London, 773-802. https://doi.org/10.5772/55749
20. Ghasiya, P., & Sasahara, K. (2022). Rapid Sharing of Islamophobic Hate on Facebook: The Case of the Tablighi Jamaat Controversy. Social Media + Society, 8(4), 205630512211291. https://doi.org/10.1177/2056305122112 9151
21. Asians, E. C., Free-, P. R., Approach, P., & Extremism, A. V. (n.d.). Media and Social Media Analysis on Religious Freedom and Violent Extremism in Central Asia : Cases of Kazakhstan , Tajikistan.
22. Strategic Communications. (2022). Hate speech poisons societies and fuels conflicts. https://www.eeas.europa.eu/eeas/hatespeech-poisons-societies-and-fuelsconflicts_en
23. Association, American library. (2017). Hate Speech and Hate Crime", American Library Association. https://doi.org/aa35c1c7-f3aa-4b07- 964f-30dcf85a503c
24. Hate, N. O. (2020). Racism, Intolerance, Hate Speech. Kiper, J. (2023). Religious Hate Propaganda: Dangerous Accusations and the Meaning of Religious Persecution in Light of the Cognitive Science of Religion. In Religions (Vol. 14, Issue 2). https://doi.org/10.3390/rel14020185
25. Catherine Runswick-Cole (2014) 'Us' and 'them': the limits and possibilities of a 'politics of neurodiversity' in neoliberal times, Disability & Society, 29:7, 1117-1129, DOI: 10.1080/09687599.2014.910107
26. Saha K, Torous J, Ernala SK, Rizuto C, Stafford A, De Choudhury M. A computational study of mental health awareness campaigns on social media. Transl Behav Med. 2019 Nov 25;9(6):1197-1207. doi: 10.1093/tbm/ibz028. PMID: 30834942; PMCID: PMC6875652.
27. Chetty, Naganna,Alathur, S. (2018). Hate speech review in the context of online social networks. Aggression and Violent Behavior, 40(March 2017), 108–118. https://doi.org/10.1016/j.avb.2018.05.0 03
28. Njagi, Joan & Ireri, Anthony & Njagi, Eliud & Augustine, Afullo & Ngugi, Mathew & Ni, Karugu. (2013). Assessment of Knowledge, Attitude and Perceptions of Village Residents on the Health Risks Posed by Kadhodeki Dumpsite in Nairobi, Kenya. World Environment. 6. 155-160. 10.5923/j.env.20130305.02.
29. MacAvaney S, Yao HR, Yang E, Russell K, Goharian N, Frieder O. Hate speech detection: Challenges and solutions. PLoS One. 2019 Aug 20;14(8):e0221152. doi: 10.1371/journal.pone.0221152. PMID: 31430308; PMCID: PMC6701757.
30. Mahesh, B. (2020) Machine Learning Algorithms—A Review. International Journal of Science and Research, 9, 381-386.
31. Alzubaidi, L., Zhang, J., Humaidi, A. J., AlDujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., AlAmidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. In Journal of Big Data (Vol. 8, Issue 1). Springer International Publishing. https://doi.org/10.1186/s40537-021- 00444-8
32. Sarker, I.H. (2021) Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science, 2, Article No. 160. https://doi.org/10.1007/s42979-021-00592-x
33. Yimam, S. M., Ayele, A. A., & Biemann, C. (2019). Analysis of the ethiopic twitter dataset for abusive speech in amharic. ArXiv, 1–5.
34. Thomas, Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017, 512–515.
35. Tesfaye, S. G., & Tune, K. K. (2020). Automated Amharic Hate Speech Posts and Comments Detection Model Using Recurrent Neural Network. Research Square. https://www.researchsquare.com/article/rs114533/latest?utm_source=researcher_app&utm_medium=referral&utm_cam paign=RESR_MRKT_Researcher_inbound
36. Das AK, Islam MN, Billah MM, Sarker A (2021) COVID-19 pandemic and healthcare solid waste management strategy – a mini-review. Sci Total Environ 778:146220. https://doi.org/10.1016/j.scitotenv.2021.146220
37. Ababu, T. M., & Woldeyohannis, M. M. (2022). Afaan Oromo Hate Speech Detection and Classification on Social Media. Proceedings of the Thirteenth Language Resources and Evaluation Conference, June, 6612–6619. https://aclanthology.org/2022.lrec1.712

Open Access
Print ISSN:2992-6106
Online ISSN:2992-5789

CITE AS: Ezeaku Florence Uzoaji (2024). Analyzing the Application of Machine Learning in Detecting Hate Speech: A Review. NEWPORT INTERNATIONAL JOURNAL OF CURRENT RESEARCH IN HUMANITIES AND SOCIAL SCIENCES 4(2):18-24. https://doi.org/10.59298/NIJCRHSS/2024/4.2.1823