INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 04, April 2024, pp: 2616-2619

EMPOWERING ONLINE SAFETY WITH ADVANCED COMMENT TOXICITY DETECTION USING MACHINE LEARNING

Mrs. Rasika Rewatkar[1], Ms. Ragini Thakrele[2], Ms. Rachana Madankar[3], Mr. Satvik Raut[4], Ms. Kumud Bankar[5], Ms. Shailee Goverdhan[6]

[1,2,3,4,5,6]Department of Information Technology Kavikulguru Institute of Technology and Science, Nagpur, India.

ABSTRACT

Online communication platforms are meant for connection and collaboration, but some misuse them to spread hate and abuse. With the surge in online interactions, manually filtering hateful content becomes nearly impossible. To combat this, we aim to use text mining and logistic regression, a type of machine learning algorithm, for binary classification to identify and filter out hate speech. Our model achieves high precision, recall, and accuracy in classifying comments as toxic or non-toxic, making social media cleaner and safer.

Key words- Hate Speech Detection, Logistic regression ,Deep Learning, Text Mining, Toxicity Classification..

1. INTRODUCTION

Online platforms are where people chat, share, and connect. But sometimes, some people use them to say nasty things. This can hurt others. With so many people online, it's hard to stop all the mean stuff manually. So, we need a way to find and remove hateful comments to make the internet a nicer place.This project aims to do just that by using computer techniques to spot hate speech. The goal is to build a smart system that can tell if a comment is mean or not. And guess what? Our system is pretty good at it! It can spot mean comments with high accuracy.

Benefits of a Advanced Comment Toxicity Detection:

Empowering online safety through advanced comment toxicity detection using machine learning offers several significant benefits:

1. Early Detection and Prevention of Harmful Content: Machine learning algorithms can analyze large volumes of user-generated content in real-time, enabling platforms to swiftly identify and mitigate toxic comments before they escalate into harmful situations. By proactively detecting toxicity, platforms can create safer and more supportive online environments for users.
2. Improved User Experience: Online platforms are often plagued by toxic behavior, which can deter users from engaging in discussions or sharing their opinions freely. By implementing effective comment toxicity detection mechanisms, platforms can foster a more positive and inclusive user experience, encouraging active participation and constructive dialogue.
3. Reduction of Cyberbullying and Harassment: Cyberbullying and harassment are pervasive issues in online communities, causing emotional distress and psychological harm to victims. Machine learning-based toxicity detection systems can flag and remove abusive comments, thereby reducing the prevalence of cyberbullying and creating a safer space for users to interact.
4. Enhanced Content Moderation Efficiency: Manual content moderation is labor-intensive and often inadequate for handling the sheer volume of user-generated content on social media platforms and online forums. Automated toxicity detection using machine learning can significantly reduce the burden on human moderators, allowing them to focus their efforts on more nuanced cases and strategic interventions.

Challenges of advanced comment toxicity detection: Implementing a comment toxicity detection can be challenging, despite its many benefits. Here are some of the key challenges associated with such a system:

1. Ambiguity and Subjectivity of Toxicity: Determining what constitutes toxic behavior in online comments can be highly subjective and context-dependent. Machine learning models may struggle to accurately discern between harmless banter, legitimate criticism, and genuinely harmful content, leading to false positives or negatives in toxicity detection.
2. Adversarial Attacks and Evasion Techniques: Malicious users may actively seek to circumvent comment toxicity detection systems by employing evasion techniques, such as subtle language manipulation, code-switching, or using coded language. Adversarial attacks can undermine the effectiveness of machine learning models and necessitate constant adaptation and retraining to maintain efficacy.

3. Privacy and Data Security Concerns: Implementing advanced comment toxicity detection systems involves analyzing large volumes of user-generated content, raising concerns about user privacy and data security. Ensuring compliance with privacy regulations, safeguarding sensitive user information, and mitigating the risk of data breaches are critical considerations in the development and deployment of such systems.

Opportunities for Society Communities:

1. Creating Safer Digital Spaces: Advanced comment toxicity detection using machine learning can help create safer digital spaces where individuals feel empowered to express themselves without fear of harassment or abuse. By effectively identifying and mitigating toxic comments, online platforms can foster a culture of respect and civility, enhancing the overall user experience.
2. Empowering Victims of Online Abuse: Victims of online harassment and cyberbullying often feel powerless in the face of toxic behaviour. Advanced comment toxicity detection can empower victims by providing them with tools to report and address abusive comments, fostering a sense of agency and control over their online experiences.
3. Supporting Mental Health and Well-being: Exposure to toxic content online can have detrimental effects on mental health and well-being, leading to increased stress, anxiety, and depression. By mitigating toxicity through machine learning-based moderation, society can help protect the mental health of individuals and promote a safer and more supportive online environment for all users.

2. IDENTIFY, RESEARCH AND COLLECT IDEA

2.1 Functional Requirements

Creating an empowering online safety with advanced comment toxicity detection involves defining functional requirements to specify can encompass various aspects of the system's functionality. Below are some key functional requirements for such comment toxicity detection:

1. Real-time Comment Monitoring: The system should be capable of monitoring user-generated comments in real-time across various online platforms, including social media, forums, and messaging apps.
2. Toxicity Detection: The system should accurately detect toxic comments using machine learning algorithms, identifying various forms of toxicity such as hate speech, harassment, threats, and abusive language.
3. Multilingual Support: The system should support multiple languages to effectively detect toxicity in comments across diverse linguistic communities and cultural contexts.
4. Scalability: The system should be scalable to handle large volumes of user-generated content, accommodating the dynamic nature of online platforms and the ever-growing scale of digital interactions.
5. Customization and Adaptability: The system should allow for customization and adaptation to specific platform requirements, user preferences, and moderation policies, ensuring flexibility and relevance in different contexts.
6. The reporting and analytics aspects can be discussed in detail, highlighting their significance and potential impact. Here's an outline of reporting and analytics considerations:
7. Data Collection and Aggregation: Describe how data on user-generated comments and toxicity detection outcomes are collected and aggregated from various online platforms. Discuss the importance of data integrity, consistency, and privacy protection in the data collection process.
8. Visualization and Dashboarding: Discuss the use of data visualization techniques and dashboarding tools to present reporting and analytics insights in a user-friendly and actionable format. Consider the design principles for effective visualization, such as clarity, simplicity, and interactivity.
9. Trend Analysis and Pattern Recognition: Explore how reporting and analytics tools can facilitate trend analysis and pattern recognition to identify emerging patterns of toxic behavior and evolving moderation challenges. Discuss the importance of proactive monitoring and intervention to address emerging threats promptly.

3. WRITE DOWN YOUR STUDIES AND FINDINGS

- Importance of Waste Management: Waste management is crucial for environmental sustainability and resource conservation. Implementing intelligent systems can enhance traditional waste management practices by automating certain processes and improving efficiency.
- Role of Machine Learning: Machine learning plays a vital role in developing intelligent waste classification systems. By training algorithms on large datasets of images or sensor data, these systems can accurately classify different types of waste materials.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 04, April 2024, pp: 2616-2619

- Data Collection and Preprocessing: Gathering diverse and comprehensive datasets is essential for training effective machine learning models. Data preprocessing techniques such as cleaning, normalization, and augmentation help improve the quality of the dataset and enhance model performance.
- Feature Extraction and Selection: Identifying relevant features from the data is critical for building robust classification models. Feature extraction techniques such as deep learning-based feature learning or handcrafted feature engineering can capture important characteristics of waste materials.
- Model Selection and Evaluation: Choosing appropriate machine learning algorithms and architectures is crucial for achieving high classification accuracy. Techniques such as convolutional neural networks (CNNs), support vector machines (SVMs), or ensemble methods can be employed and evaluated based on metrics like precision, recall, and F1-score.
- Real-Time Implementation and Deployment: Integrating the intelligent waste classification system into real-world environments requires considerations for scalability, computational efficiency, and practical usability. Deployment on edge devices or cloud platforms can enable real-time monitoring and decision-making in waste management processes.
- Continuous Improvement and Adaptation: Continuous monitoring and feedback mechanisms are essential for iteratively improving the performance of the classification system. Techniques like transfer learning or online learning can adapt the model to evolving waste compositions and environmental conditions.
- Socio-Economic and Environmental Impacts: Assessing the socio-economic and environmental impacts of implementing intelligent waste classification systems is essential for evaluating their overall effectiveness. These systems should contribute to reducing landfill waste, promoting recycling initiatives, and fostering a circular economy.

4. GET PEER REVIEWED

- Submission to Journals or Conferences: After completing your research study or paper, you would typically submit it to reputable academic journals or conferences in your field.
- Evaluation by Experts: Upon submission, the manuscript undergoes evaluation by peer reviewers who are experts in the subject matter. They assess the quality, originality, methodology, and significance of the research.
- Feedback and Revisions: Peer reviewers provide feedback and constructive criticism to the authors, highlighting strengths and weaknesses of the work. Authors may need to revise their manuscript based on the reviewers' comments.
- Decision by Editorial Board: Based on the peer reviewers' recommendations and the overall quality of the manuscript, the journal's editorial board makes a decision on whether to accept, reject, or request revisions to the paper.
- Publication: If accepted, the research paper is published in the journal or presented at the conference, contributing to the scholarly literature in the field.
- Continuous Improvement: The peer review process helps ensure the integrity and credibility of academic research. It also facilitates knowledge exchange and fosters continuous improvement in research methodologies and practices

5. SYSTEM ARCHITECTURE

The system architecture gives an overview of the working of the system.

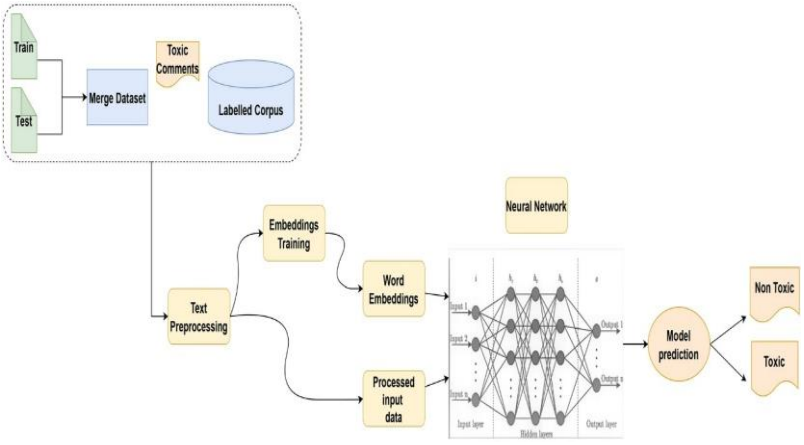

Fig 3. SYSTEM ARCHITECTURE

6. CONCLUSION

In conclusion, the development and implementation of an Intelligent Waste Classification System using Machine Learning present a promising avenue for addressing the challenges of modern waste management and advancing sustainable resource management practices. By leveraging machine learning algorithms and sensor technologies, such systems offer the potential to automate waste sorting processes, enhance recycling efforts, and minimize environmental impacts associated with improper waste disposal. While there are challenges to overcome, including data collection, model optimization, and real-world deployment, the benefits of intelligent waste classification systems in promoting environmental sustainability and resource conservation are substantial. Continued research and collaboration across disciplines are crucial for further refining these systems and realizing their full potential in creating a cleaner, greener future.

7. REFERENCES

[1] Gambäck, B., Sikdar, U.: Using Convolutional Neural Networks To Classify Hate-Speech. In Proceedings Of The First Workshop On Abusive Language Online, Pages 85–90.Association For Computational Linguistics (2017)

[2] Chu, T., Jue, K., Wang, M.: Comment Abuse Classification With Deep Learning. Stanford University (2017)

[3] Zhang, Z., Robinson, D., Tepper, J.: Detecting Hate Speech On Twitter Using A Convolution-Gru Based Deep Neural Network. In Proceedings Of The 15th Extended Semantic Web Conference, Eswc18, Pages 745–760. Springer (2018)