IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 11, Issue 1, January 2024

Impact Factor: 7.394

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

Detection of Cyber Bullying on Social Media

Mrs.R.Kavitha, Mr.J.Sakthivel, Mr.A.Pugazharasan, Mr.Z.Mohamed zawith,

Assistant Professor, Department of Cyber Security, Muthayammal Engineering College, Rasipuram, India

Student, Department of Cyber Security, Muthayammal Engineering College, Rasipuram, India

Student, Department of Cyber Security, Muthayammal Engineering College, Rasipuram, India

Student, Department of Cyber Security, Muthayammal Engineering College, Rasipuram, India

ABSTRACT: While social media offer great communication opportunities, they also increase the vulnerability of young people to threatening situations online. Recent studies report that cyber-bullying constitutes a growing problem among youngsters. Implement Semantic approach with classification algorithm using Deep learning algorithm to classify the messages whether is positive or not Block friends by predefined threshold value

As the ubiquity of social media platforms continues to rise, so does the prevalence of cyberbullying, presenting a pressing challenge for maintaining a safe online environment. This research explores the development and implementation of robust systems for the detection of cyberbullying on social media platforms. The study emphasizes the significance of proactive measures to identify and mitigate instances of cyberbullying, aiming to create a safer and more inclusive digital space.

The proposed system employs advanced natural language processing (NLP) techniques and machine learning algorithms to analyze textual content on social media platforms. By leveraging sentiment analysis, linguistic patterns, and contextual cues, the system aims to identify potentially harmful or abusive language indicative of cyberbullying. Additionally, the incorporation of user behavior analysis enhances the system's ability to recognize patterns consistent with cyberbullying dynamics.

I. INTRODUCTION

Text analysis, a vital field within natural language processing (NLP), plays a crucial role in extracting meaningful insights and valuable information from extensive amounts of textual data.[1] As digital communication and the utilization of social media platforms continue to grow exponentially, the importance of text analysis in understanding human behavior, sentiment, and discourse patterns has significantly increased. This surge in demand for text analysis techniques is driven by the availability of massive textual data and the development of advanced machine learning algorithms. These techniques find applications in various domains, such as sentiment analysis, topic modeling, information retrieval, and more. Through the application of these methods, researchers and practitioners can uncover valuable insights, patterns, and trends embedded within textual data.

In today’s digital society, cyberbullying has emerged as a prevalent issue, referring to the intentional and repetitive use of digital communication platforms to harass, intimidate, or harm individuals. Cyberbullying encompasses a wide range of harmful behaviors, including the dissemination of rumors, sharing explicit or defamatory content, sending abusive messages, and engaging in online hate speech. The proliferation of digital platforms, such as social media, online forums, and messaging applications, has provided individuals with unprecedented means of communication and expression. However, it has also created breeding grounds for cyberbullying, enabling perpetrators to target their victims anonymously or under false identities, exacerbating the detrimental effects of their actions. The negative impact of cyberbullying on individuals, particularly their mental health, social interactions, and overall well-being, cannot be overstated. Victims often experience heightened levels of stress, anxiety, depression, and decreased self-esteem.

The persistent nature of online harassment can lead to social isolation, strained relationships, and hindered academic or professional performance. In severe cases, cyberbullying has even been linked to self-harm and suicidal ideation among victims. The purpose of this research paper is to contribute to the existing body of knowledge on cyberbullying by proposing a novel approach to its detection.By leveraging the power of federated learning and text analysis techniques, we aim to develop a robust and privacy-preserving framework that can effectively identify and combat instances of cyberbullying while upholding user privacy. Our research aligns with the overarching objective of creating safer online environments and promoting the well-being and mental health of individuals in the digital era. Bycombining federated learning, which allows for collaborative model training while preserving data privacy, with advanced text analysis techniques, we seek to empower ions and individuals to collectively address the pressing issue of cyberbullying.

II. SYSTEM ANALYSIS

- In collaborative filtering information will be selected on the basis of user's preferences, actions, predicts, likes, and dislikes.
- In policy based filtering system users filtering ability is represented to filter wallmessages according to filtering criteria of the user.
- Machine learning algorithms including Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbours (KNN), Naïve Bayes (NB), Decision Trees (DT) to provide rules based on their relationships to reduce the complexity of classification.
- Advanced machine learning algorithms are employed to analyze patterns of behavior and language to identify potential instances of cyberbullying. These algorithms can adapt and improve over time as they are exposed to more data.
- Monitoring user behavior, such as the frequency of interactions, reporting, and blocking, can also be part of a system for detecting cyberbullying. Unusual patterns may trigger alerts for further investigation.
- NLP techniques are used to analyze the context and sentiment of messages. This helps in distinguishing between casual conversation and potentially harmful content.
- Some platforms employ community moderation, where users collectively enforce community guidelines by reporting and flagging inappropriate content. This can be supplemented with automated systems.

- Most social media platforms have reporting mechanisms that allow users to report instances of cyberbullying.

Limitations:

- There is no approach for filter unwanted messages in social networks.
- Current systems may struggle with understanding the context of a conversation, making it challenging to distinguish between harmless banter and actual instances of cyberbullying.
- Different cultures and languages express concepts differently, making it difficult for systems to account for the nuances of various linguistic styles and cultural norms.
- As cyberbullying is not limited to text, the detection of harmful content in images, videos, and audio poses additional challenges. Analyzing multimedia content for contextual understanding and identifying abusive elements can be complex.
- Can't deal with posts that are associated with images.
- Difficult analyze short text tags.
- Automatic blocking can't be implemented.

Proposed System:

- Implement classification algorithm based on natural language processing toanalyze good and bad comments.
- Eliminate the stop words, stemming words and extract key terms based on textmining approach.
- Classification can be include back propagation neural network algorithm to labelcomments.
- Implement a keyword filtering system to flag potentially offensive words orphrases commonly associated with cyberbullying.
- Utilize NLP algorithms to analyze the context and sentiment of messages,allowing for a more nuanced understanding of language and conversation.
- Train machine learning models to identify patterns of behavior associated with cyberbullying. Use labeled datasets to teach the system to recognize both explicitand subtle forms of harassment.
- SMS alert at the time of posting negative comments in social network page.
- Develop mechanisms for dynamically updating keyword lists and machine learning models to adapt to the evolving nature of language and emerging cyberbullying trends.
- Block the friends who continuously post the negative comments.

Excepted Merits:

- Short texts are classified based neural network approach.
- And also blocked malicious users.

- System uses deep learning classifier to enforce customizable content dependent rules.
- Early detection allows for prompt intervention, helping to prevent the escalationof cyberbullying incidents. Timely action can mitigate harm and discourage further abusive behavior.
- Users are more likely to feel safe and secure on social media platforms when there are effective cyberbullying detection mechanisms in place. This can lead to a positive user experience and increased user retention.
- A safer online environment fosters positive community building. Users are more likely to engage in constructive discussions and share their thoughts and experiences when they feel protected from harassment.

ALGORITHM

TEXT MINING ALGORITHM

- Tokenize text-based reviews as single terms
- Analyze unigrams, bigrams, and n-grams
- Remove stop words, analyze stemming words, and remove special characters
- Finally, extract key phrases
- Analyze extended words that can be substituted with right words function INITBPNNMODEL (θ, [n1–5])
- layerType = [convolution, max-pooling, fully-connected, fully-connected];
- layerActivation = [tanh(2), max(),softmax()]
- model = new Model();
- for$i$=1 to 4 do
- layer = new Layer();
- layer.type = layerType[i];
- layer.inputSize = ni
- layer.neurons = new Neuron [ni+1];
- layer.params = θi;
- model.addLayer(layer);
- end for
- return model;
- end function

SOFTWARE DESCRIPTION:

Hardware Requirements
System : Intel processor
Hard disk : 100MB
Monitor : 14 Inch Color Monitor
RAM : 1 GB

Software Requirements
Operating System : Windows Front End : PYTHON
Back End : MYSQL
Tool : PYCHARM

Software Description

PyCharm provides a powerful code editor with features like syntax highlighting, code completion, and error highlighting, making it easier for developers to write clean and error-free Python code.

The IDE offers intelligent code assistance, including code completion suggestions, quick fixes, and code navigation tools. This helps developers increase their productivity and write code more efficiently.

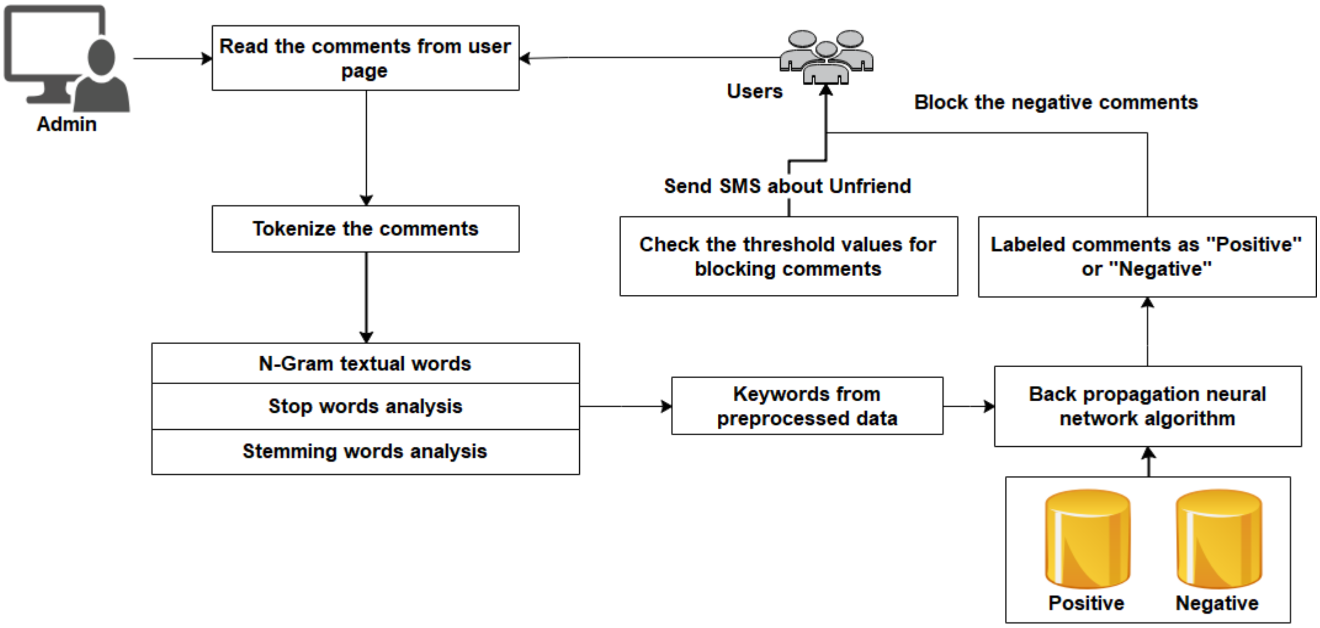
PyCharm comes with a built-in debugger that allows developers to easily debug their Python code. It supports features like breakpoints, stepping through code, and variable inspection.

It has robust support for writing and running tests, including integration with popular testing frameworks like pytest and unittest. Test results are displayed within the IDE, making it easy to identify and fix issues.

PyCharm integrates with version control systems such as Git, Mercurial, and Subversion. This allows developers to manage their code repositories directly within the IDE.

PyCharm is not limited to Python-only development. It provides support for web development technologies, including HTML, CSS, and JavaScript. It also supports popular web frameworks like Django and Flask.
PyCharm assists developers in managing virtual environments, which are isolated Python environments for projects. It helps create, configure, and switchbetween virtual environments seamlessly.
The Professional edition includes additional features such as scientific tools, advanced web development support, database tools, and support for popular frameworks like Angular, React, and Vue.js.

System Architecture:


Admin
Read the comments from user page
Users
Block the negative comments
Send SMS about Unfriend
Tokenize the comments
Check the threshold values for blocking comments
Labeled comments as "Positive" or "Negative"
N-Gram textual words
Stop words analysis
Stemming words analysis
Keywords from preprocessed data
Back propagation neural network algorithm
Positive
Negative

III. CONCLUSION

- In this project, we have presented a system to filter undesired messages from OSN walls.
- The system exploits a DL soft classifier to enforce customizable content dependent filtered rules system.
- The major efforts in building a robust short text classifier are concentrated in the extraction and selection of a set of characterizing and discriminant features.
- Moreover, the flexibility of the system in terms of filtering options is enhanced through the management of BLs.

REFERENCES

1. Roy, Pradeep Kumar, et al. "A framework for hate speech detection using deep convolutional neural network." IEEE Access 8 (2020): 204951-204962.
2. Aluru, Sai Saketh, et al. "Deep learning models for multilingual hate speech detection." arXiv preprint arXiv:2004.06465 (2020).
3. Mullah, Nanlir Sallau, and Wan Mohd Nazmee Wan Zainon. "Advances in machine learning algorithms for hate speech detection in social media: a review." IEEE Access 9 (2021): 88364-88376.
4. Cao, Rui, Roy Ka-Wei Lee, and Tuan-Anh Hoang. "DeepHate: Hate speech detection via multi-faceted text representations." Proceedings of the 12th ACMConference on Web Science. 2020
5. Khan, Shakir, et al. "BiCHAT: BiLSTM with deep CNN and hierarchical+ attention for hate speech detection." Journal of King Saud University-Computer and Information Sciences 34.7 (2022): 4335-4344.

IJARETY

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 Impact Factor: 7.394

www.ijarety.in editor.ijarety@gmail.com