www.ijprems.com
editor@ijprems.com

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

Vol. 03, Issue 05, May 2023, pp : 1165-1168

e-ISSN : 2583-1062

Impact Factor : 5.725

SOCIAL MEDIA CONTENT FILTERING

Madhuri Palve[1], Shubhangi Kangane[2], Shubhangi Waghmare[3], Prof. S. M. Rokade[4]

[1,2,3]BE Student, Department of Computer Engineering, Sir Visvesvaraya Institute of Technology, Nashik

[4]Professor, Department of Computer Engineering, Sir Visvesvaraya Institute of Technology, Nashik

DOI: https://www.doi.org/10.58257/IJPREMS31447

ABSTRACT

Due to the recent explosion of social media usage, communication has improved significantly. People use them to keep in touch with old acquaintances, make new ones, update any significant events in their lives, etc.

These media are heavily populated with spam messages as a result of their wide reach and rising popularity.

The majority of spam messages are distributed at random to many recipients by dishonest advertisers and criminals who want to direct you to phishing websites.

In order to identify spam in short text messages and to identify spam photos, we are applying a variety of machine learning approaches.

Keywords: Social, popularity, spam, approach, communication, Photos.

1. INTRODUCTION

Social media content filtering is the practice of keeping an eye on and policing the information shared on social networking sites to make sure it complies with accepted practices, laws, and regulations. In order to review and filter out content that might be deemed improper, offensive, dangerous, or in violation of platform regulations, it makes use of both automated algorithms and human moderation.

The massive amount of user-generated information and the exponential expansion of social media platforms have made content filtering essential for preserving a secure and encouraging online environment. The goal of content filtering systems is to strike a compromise between preserving the right to free speech and halting the spread of dangerous or unsuitable content.

Filtering social media content often entails a number of processes. The content is first automatically scanned utilizing machine learning, computer vision, and natural language processing (NLP) techniques. These algorithms scan the text, pictures, and videos posted on social networking sites to find potentially offensive or illegal material.

The algorithms use a number of signals and features, such as sentiment analysis, image recognition, keyword matching, and context analysis, to make filtering judgements. They are able to recognize and flag material that, among other things, contains hate speech, harassment, violence, nudity, violent imagery, or copyright violations.

Following the initial automated filtering, the flagged content is next examined by human moderators, who ultimately decide whether it should be deleted or kept. Human moderators exercise discretion while considering the platform's rules, regional laws, and cultural sensitivity.

The process of content screening on social media networks is intricate and constantly changing. Platform providers make ongoing research and development investments to raise the precision and efficiency of their filtering techniques. They work to fix issues like false positives (material that is incorrectly recognized as being against the rules) and false negatives (stuff that breaks the rules but isn't caught).

2. METHODOLOGY

- Spam might or might not be damaging to the recipient. It might be anything from a lighthearted text message to a dangerous virus that could ruin your entire computer, or a malicious code designed to steal all the data on your computer.
- The proliferation of spam initially began with email, but as the Internet and social media became more popular, it began to spread like an epidemic.
- Due to the substantial reduction in spamming caused by growing email spam awareness, traditional spammers are now embracing mobile and Internet technologies as spam mediums.
- Therefore, we are blocking spam messages, URLs, and photos on our system.
- Our system will block any attempts by a user to send offensive language, inappropriate content, photos, or phishing links, and the user will be forced to log out of the system.
- The other user will not get the spam mail.

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

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 05, May 2023, pp : 1165-1168

e-ISSN : 2583-1062

Impact Factor : 5.725

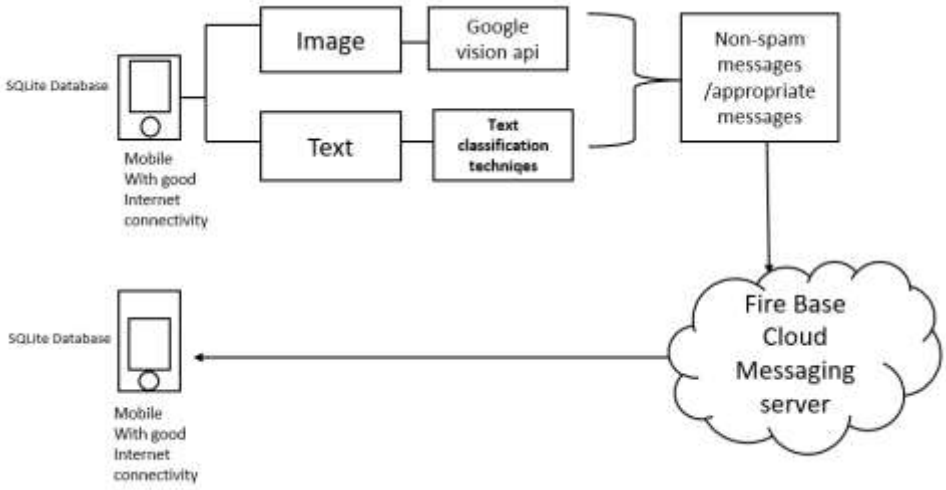

Figure 1: System Architecture

3. MODELING AND ANALYSIS

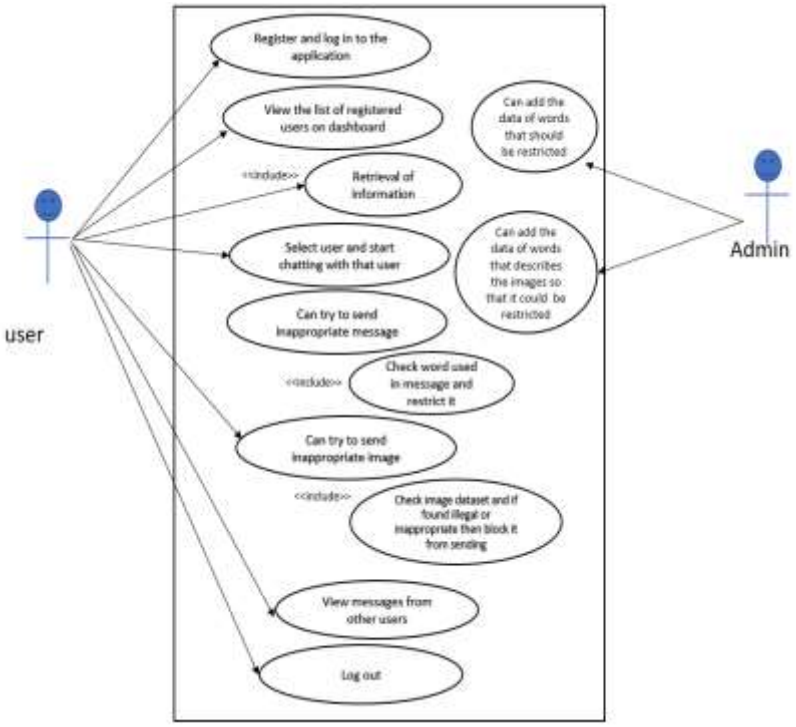

Figure 2: Use-case Diagram

Planning and Prototyping:

In proposed system, the waterfall model has essentially been used.

1. **Requirements**: As previously noted, a preliminary study of all requirements, including those for hardware, software, and non-functional components, is carried out.
2. **Design:** Project constraints, risk analysis, technology analysis, schedule analysis, and other tasks are completed. then all the system architecture, database, and user flow components are located.
3. **Implementation:** In this stage, we began writing code in accordance with the specifications and system design.

Validation:

Testing for validation will be done in the final step.We'll run a variety of tests, including the following:

1. **Functional test**: This will make sure that no features have been left out. The accuracy of all the features described during the initial phase will be tested.
2. **System testing:** This is carried out once the entire project has been completed. To confirm that no errors have been introduced, all of the queries are run once more.
3. **Logical testing:** To ensure thorough application testing, the system will be examined using various test data.

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

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 05, May 2023, pp : 1165-1168

e-ISSN : 2583-1062

Impact Factor : 5.725

4. RESULTS AND DISCUSSION

Filtering social media content can have both beneficial and negative effects. Users may benefit from social media news filtering because it uses algorithmic filtering to provide just specific information, which lessens anxiety and the cognitive load of processing news. Social media filtering, however, may sometimes be detrimental, exposing users to unsavory political debates and forming filter bubbles that stifle exposure to opposing views. Using machine learning approaches, content filtering on social media platforms is possible and may include auto-blocking features. Legal and technological analysis are topics related to the effects of algorithms for online content screening or moderation. The effectiveness and potential downsides of social media content filtering on various platforms require further study.

Figure 3: Login

Figure 4: Welcome page for App

Figure 5: OTP Sent

Figure 6: Verify OTP

Figure 7: Select contact

Figure 8: Select another contact.

5. CONCLUSION

In conclusion, real-time object detection using machine learning has the potential to greatly improve the quality of life for visually challenged people. By leveraging technologies such as computer vision and deep learning, it is possible to develop algorithms that can recognize and describe the visual world in real-time, allowing users to navigate and interact with their environment more easily.

There are already several promising initiatives underway to develop assistive technologies based on real-time object detection, such as wearable devices that can identify objects and provide audio feedback to the user. As these technologies continue to evolve and improve, we can expect to see even more advanced and sophisticated applications in the near future.

Overall, the use of machine learning for real-time object detection is a promising and exciting area of research, with the potential to transform the lives of visually challenged people and help them to live more independently and confidently.

6. REFERENCES

[1] L. F. Cranor, & B. A. LaMacchia, “Spam!”, Communications of the ACM, Vol 41, Issue 8, pp.74-83. 1998.

[2] G. Jain, & M. Sharma, “Social Media: A Review”, In Information Systems Design and Intelligent Applications, Springer India, pp. 387-395. 2016

[3] M. Nelson, “Spam Control: Problems and Opportunities”, Ferris Research, India, pp.23-82, 2003. [4] G.V. Cormack, J.M.G. Hidalgo, E.P. Sánz, “Feature engineering for mobile (SMS) spam filtering”, In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, US, pp. 871-872, 2007.

[4] Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. Proceedings of the International AAAI Conference on Web and Social Media (ICWSM), 512-515.

[5] Dadvar, M., & Trieschnigg, D. (2013). Towards context-aware filtering of hate speech content on Twitter. Proceedings of the International Conference on Weblogs and Social Media (ICWSM), 573-576.

[6] Waseem, Z., & Hovy, D. (2016). Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter. Proceedings of the NAACL Student Research Workshop, 88-93.

[7] Ribeiro, M. H., Araújo, M., Gonçalves, P., & Benevenuto, F. (2018). SentiBench—a benchmark comparison of state-of-the-practice sentiment analysis methods. EPJ Data Science, 7(1), 45.

[8] Zubiaga, A., & Ji, H. (2014). Tweet, but verify: epistemic study of information verification on Twitter. Proceedings of the International Conference on Weblogs and Social Media (ICWSM), 641-650.

[9] Mathew, B. K., Bhattacharyya, P., & Ghosh, S. (2019). Overview of multimedia data processing and analytics in social media: Extraction, summarization, and content filtering. Multimedia Tools and Applications, 78(2), 1573-1611.

[10] Park, M., & Fink, E. (2019). Examining the effectiveness of social media content moderation: Evidence from Reddit. Proceedings of the ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW), 442-453.