Content Moderation Technique using hybrid approch

Prof. Avinash Palave[1], Sanafarin Mulla[2]
[1, 2] Department of Computer Engineering
[1, 2] TCOER, Savitribai Phule Pune University, Kondhwa, Pune-48, Maharashtra, India.

Abstract- *The web is growing, and the extent of active user participation on the web is growing with it as well. There are many social media sites receives millions of posts and comments each day. Many of these posts are insightful or informative, but there is a large quantity of abuse as well. According to a recently released study, as much as 95% of user-generated posts on Web sites are spam or malicious. To help combat the threat of unchecked UGC, it's important to develop some moderation technique to manage the influx of user-generated content as business grows which can helps business to protect their investment over the long term. To improve the scalability and efficiency this paper proposed the implementation in mix moderation technique.*

Keywords- Data Filtration, Content Moderation, Cyber bullying, Online Content

I. INTRODUCTION

In today's world, while the Internet is a wonderful thing, the anonymity it provides can sometimes bring out the worst in people. Most web-sites that display content, also allow user comments and discussions, and attract a huge volume of user posts for instance. But bad or abusive user-generated content like comments, post, photos, videos among others on your website, can seriously affect the image of a company and tarnish its reputation which, obviously, is bad for business. With this paper our objective is to provide real-time moderation around the clock to avoid delays in pulling inappropriate content and allowing beneficial content to post—naturally giving your content marketing a boost and at the same time offer the benefits of Minimization of inappropriate content, Increased credibility, Spam control, Cost and time savings.

II. EXISTING SYSTEM

Majority of websites accepting comments, images, and videos from users have a review mechanism in place to filter out content that violates their terms of service. Some site owners choose to review submissions after they've already gone live while Others place user generated content into a moderation queue for manual review

No matter which approach you take, one thing is for certain: effective content moderation is critical for your online business. Many businesses prefer to employ a team of human reviewers to filter through user generated content. As the volume of user-generated content (UGC) increases, a solitary trusted moderator cannot single-handedly deal with the problem of identifying bad content. Studies have shown that prolonged exposure to violent, pornographic, and disturbing material can be damaging for reviewers.

III. PROBLEM STATEMENT

Moderating such huge amount of data will not be in scope of human or manual moderations. It might put humans at higher risk for developing depression and anxiety; they may also have anger problems and trouble maintaining healthy relationships. On the other hand we can't rely completely on automatic or machine moderation where most of may not filtered accurately or may some good content get blocked.

IV. PROPOSED SYSTEM

This new system, we choose to implement an mix automated solution to remove some of the stress from human review team. Adding an hybrid automated program to help support human reviewers can help you protect your employees, streamline the moderation process, and drastically reduce your business' overall costs of putting large number of workforce on content moderation

The proposed system used collaborative filtering method where the content is first moderated by machine and if some keywords are matched with the inappropriate words in our database then the content will be aligned for human moderation.

End user will post content or upload file on portal. System will detect the file type based on its extension. The file which are detected as inappropriate by system will be saved in database and queued for manual review of moderator

For text file or comments: Text-based moderation will be done with the use of logic that detect inappropriate words and flag posts that may need verification. Fraudulent reviews, spams, and offensive posts will be stacked for manual moderation

For Image file: The image file will be validated using 3rd party API. If qualify the predetermined quality standards set in

the API then image will be directly posted on the portal else it will be saved in database for manual review of the moderator. Guidelines will set to make sure that the images are appropriate to the audience

For Video file: The video file will be validated same as image file using 3rd party API. It follows almost the same rules that has been implemented by most of the popular sites like YouTube, except that it can be more technical, system will perform frame by frame analysis to moderate the videos. If qualify the predetermined quality standards set in the API then image will be directly posted on the portal else it will be saved in database for manual review of the moderator. Guidelines will set to make sure that the videos are appropriate to the audience

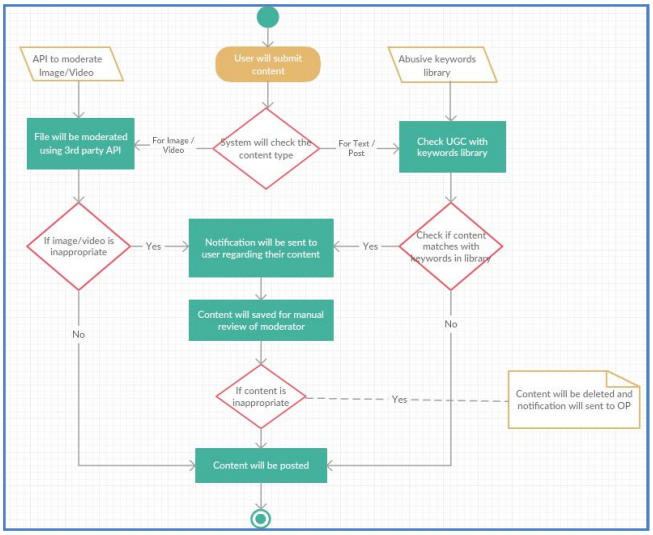

Fig.1. Content Moderation Flow

V. ANALYSIS

This application will give you real-time moderation around the clock to avoid delays in pulling inappropriate content and allowing beneficial content to post—naturally giving your content marketing a boost.

VI. CONCLUSION

Whether you run an established social networking site or you're in the process of starting up, it's important to take content moderation seriously to protect your community, business, and employees. Also Having an in-house staff member moderating all of this is not only time consuming, it's often cost prohibitive, produces an unrealistic workload and results in inefficiency by pulling valued team members onto moderation tasks . This application will give you real-time moderation around the clock to avoid delays in pulling inappropriate content and allowing beneficial content to posted with less human intervention. This application allows you to remove spam before it clutters your online presence and build your credibility in the eyes of your customers .With real-time content moderation by this system, misleading information, inappropriate content, images and videos are removed proactively to protect your portal and thus will help in minimization of inappropriate content at the same time it will helps organization to build more credibility in the eyes of consumers

ACKNOWLEDGMENT

I take this opportunity to express my profound gratitude and deep regards to guide, Prof. AVINASH .PALAVE for his exemplary guidance, cordial support, monitoring and constant encouragement throughout the course of this paper, valuable information which helped me in completing this task through various stages. The blessing, help and guidance given by him shall time to time carry me a long way in the journey of life on which I am about to embark. I also obliged to the staff members of the Department of Computer Engineering, Trinity College of Engineering & Research (TCOER) & Prof. JADHAV Department Coordinator for the valuable information provided by him .I am grateful for the cooperation I received during the paper preparation. I would also like to express my gratitude towards Prof. S. B. Chaudhari (H.O.D) and Dr. P. S. Dabeer, (Principal, TCOER), for their support in all aspects of learning. Lastly I would like to thank our colleagues and friends for their constant encouragement and help without which this hectic task would not be have been possible.

REFERENCES

[1] G. O. Young, "Synthetic structure of industrial plastics," in Plastics, 2nd ed., vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15-64

[2] G. O. Young, "Synthetic structure of industrial plastics," in Plastics, vol. 3, Polymers of Hexadromicon, J. Peters, Ed., 2nd ed. New York: McGraw-Hill, 1964, pp. 15-64.

[3] M. Abramowitz and I. A. Stegun, Eds., Handbook of Mathematical Functions (Applied Mathematics Series 55). Washington, DC: NBS, 1964, pp. 32-33.

[4] L. Stein, "Random patterns," in Computers and You, J. S. Brake, Ed. New York: Wiley, 1994, pp. 55-70.

[5] Transmission Systems for Communications, 3rd ed., Western Electric Co., Winston-Salem, NC, 1985, pp. 44–60

[6] Motorola Semiconductor Data Manual, Motorola Semiconductor Products Inc., Phoenix, AZ, 1989.

[7] RCA Receiving Tube Manual, Radio Corp. of America, Electronic Components and Devices, Harrison, NJ, Tech. Ser. RC-23, 1992.

[8] E. E. Reber et al, "Oxygen absorption in the earth's atmosphere," Aerospace Corp., Los Angeles, CA, Tech. Rep. TR-0200 (4230-46)-3, Nov. 1988

[9] J. H. Davis and J. R. Cogdell, "Calibration program for the 16-foot antenna," Elect. Eng. Res. Lab., Univ. Texas, Austin, Tech. Memo. NGL-006-69-3, Nov. 15, 1987.

[10] J. Jones, (1991, May 10), Networks (2nd ed.) [Online], Available: http://www.atm.com/info