International Journal of Computer Sciences and Mathematics Engineering Vol. 3 No. 1 (2024)
Published by: CV. HAWARI

International Journal of Computer Sciences and Mathematics Engineering
Journal homepage: www.ijecom.org

Application of Machine Learning in Detecting Dangerous Content on Facebook Social Media Applications to Protect Teenagers on Social Media

Saryulis[1], ZulFajri Dhia Ulhaq[2], Gilang Gemilang[3], Supina Batubara[4]
[1,2,3] Program Studi Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Pembangunan Panca Budi, Indonesia
[4]Program Studi Sistem Komputer, Fakultas Sains dan Teknolgi, Universitas Pembangunan Panca Budi, Indonesia

Article Info

Article history:

Received Jun 24, 2024
Revised Jun 27, 2024
Accepted Jun 30, 2024

Keywords:

Machine Learning
Social media Facebook
Malicious Content Detection
Cyber Security

ABSTRACT

This study investigates the utilization of machine learning techniques to identify harmful content on the social media platform Facebook, with a particular focus on protecting teens in their social media interactions. The rise of social media has exposed young users to a variety of risks, including cyberbullying, hate speech and inappropriate material. By developing machine learning models trained on a diverse dataset of text, images and videos shared on Facebook improves content moderation efforts to protect teens from exposure to harmful content. Through the application of natural language processing and image recognition algorithms, the model will classify content based on pre-defined categories of harmful material. It is hoped that these findings will contribute to the advancement of content moderation systems on social media platforms, and encourage a safer online environment for teenage users.

Corresponding Author:

Supina Batubara
Program Studi Sistem Komputer
Fakultas Sains dan Teknolgi
Universitas Pembangunan Panca Budi
Indonesia
Email: supinabatubara@dosen.pancabudi.ac.id

1. INTRODUCTION

Social media such as Facebook have become an integral part of the daily lives of many individuals, including teenagers. However, the ease of access and sharing of information on these platforms also carries significant risks, especially regarding harmful content that can harm users, especially teenagers who are vulnerable to negative influences.

E-ISSN 2962-4274

Harmful content on social media includes various forms, such as inappropriate images and videos, verbal or non-verbal abuse, as well as extremist propaganda and fraud. This issue raises complex challenges in efforts to protect users, especially teenagers who often do not have full awareness of these potential dangers.

One promising approach to overcome this challenge is the application of Machine Learning technology. Machine Learning offers the ability to automatically recognize and classify malicious content with a high degree of accuracy, based on patterns learned from extensive training data.

2. RESEARCH METHOD

The results of our search were a systematic literature study. Where in the literature study there are sequential steps carried out. The method in our research is identifying problems, looking for appropriate articles, selecting articles that apply machine learning to detect dangerous content on the Facebook application in order to protect teenagers in social media. The following is a diagram of our research method.

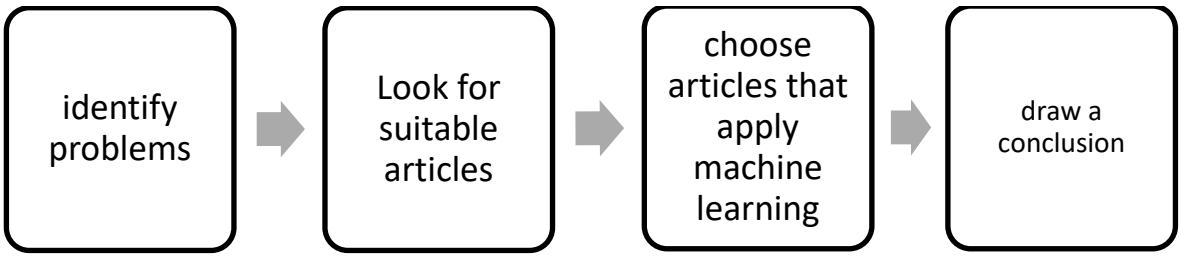

Fig 1. Research Methode

Identifying the problem that we did required the application of Machine Learning in hazard classification to detect dangerous content in the Facebook application. It is hoped that in the future the level of accuracy of the application of machine learning can help to determine whether uploaded content is dangerous or not.

3. RESULTS AND DISCUSSIONS

The application of Machine Learning in detecting dangerous content on social media applications such as Facebook has an important role in protecting teenagers on social media. The following are the results and discussion related to this topic:

Results from Applying Machine Learning:
Harmful Content Detection: Machine Learning is used to identify potentially harmful or inappropriate content such as violence, intimidation, pornography and other harmful behavior. Algorithms can scan text, images, and videos for suspicious patterns.

Fast Response: By using machine learning, Facebook can detect harmful content more quickly than if it relied solely on human moderation. This allows for quicker action to remove harmful content before it causes wider impact.

Large Scale: Social media like Facebook has millions or even billions of active users every day. Machine Learning enables automated detection that can handle huge volumes of content without requiring human intervention in each case.

E-ISSN 2962-4274

Improved Accuracy: Machine Learning algorithms are continually being improved using techniques such as deep learning to increase accuracy in recognizing new or known malicious content.

Discussion of the Application of Machine Learning:
Privacy and Ethics: Although effective, the application of Machine Learning in malicious content detection must consider user privacy issues and the ethics of data use. Facebook needs to ensure that scanning activities do not violate user privacy or raise concerns about the use of personal data.

Model Development: The process of developing Machine Learning models for malicious content detection requires large and diversified data to train the algorithms well. This includes building appropriate datasets and continuous improvement of existing models.

Collaboration with External Parties: Facebook may work with researchers, data ethicists, and advocacy groups to ensure that the Machine Learning technology it uses achieves protective objectives without compromising free speech or the security of user data.

User Education: In addition to using technology, user education and awareness about how to report and manage harmful content is also important. This can help reduce the negative impact of harmful content before Machine Learning technology can act.

4. CONCLUSION

The application of Machine Learning in detecting dangerous content on social media applications such as Facebook offers an effective solution to protect teenagers on social media. This technology enables fast and accurate detection of threatening content, such as violence, intimidation and pornography, which can reduce its negative impact on young users.

Despite this, challenges such as user privacy and ethical use of data remain major concerns. Protection of users' personal data and the balance between freedom of expression and protection from harmful content must be carefully managed in the application of this technology.

In conclusion, by continuing to improve Machine Learning models, collaborating with various parties, and increasing user awareness, Facebook can create a safer and more positive environment for teenagers and all users when interacting on this social media platform.

REFERENCES

Asnicar, F., Thomas, A. M., Passerini, A., Waldron, L., & Segata, N. (2024). Machine learning for microbiologists. *Nature Reviews Microbiology*, *22*(4), 191-205.

Hahne, F., Huber, W., Gentleman, R., Falcon, S., Gentleman, R., & Carey, V. J. (2008). Unsupervised machine learning. *Bioconductor case studies*, 137-157.

Indika, D. R., & Jovita, C. (2017). Media sosial instagram sebagai sarana promosi untuk meningkatkan minat beli konsumen. *Jurnal Bisnis Terapan*, *1*(01), 25-32.

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685-695.

E-ISSN 2962-4274

Kusuma, D. N. S. C., & Oktavianti, R. (2020). Penggunaan Aplikasi Media Sosial Berbasis Audio Visual dalam Membentuk Konsep Diri (Studi Kasus Aplikasi Tiktok). *Koneksi*, *4*(2), 372-379.

Provost, F., & Kohavi, R. (1998). On applied research in machine learning. *MACHINE LEARNING-BOSTON-*, *30*, 127-132.

Ramdani, N. S., Nugraha, H., & Hadiapurwa, A. (2021). Potensi pemanfaatan media sosial tiktok sebagai media pembelajaran dalam pembelajaran daring. *Akademika: Jurnal Teknologi Pendidikan*, *10*(02), 425-436.

Sholeh, M., Rachmawati, R. Y., & Susanti, E. (2020). Penggunaan aplikasi Canva untuk membuat konten gambar pada media sosial sebagai upaya mempromosikan hasil produk UKM. *SELAPARANG: Jurnal Pengabdian Masyarakat Berkemajuan*, *4*(1), 430-436.

Srirejeki, K. (2016). Aplikasi Media Sosial untuk Pemberdayaan USAha Mikro Kecil dan Menengah (UMKM) di Era Ekonomi Digital. *Masyarakat Telematika Dan Informasi*, *7*(1), 233770.

Wibowo, T., & Yudi, Y. (2021, September). Studi Penetrasi Aplikasi Media Sosial Tik-Tok Sebagai Media Pemasaran Digital: Studi Kasus Kota Batam. In *Conference on Business, Social Sciences and Technology (CoNeScINTech)* (Vol. 1, No. 1, pp. 662-669).