KOMTEKINFO JOURNAL
LPPM Universitas Putra Indonesia YPTK Padang
Jl. Raya Lubuk Begalung, Padang, West Sumatera, Indonesia, Zip Code: 25221
Volume: 12, Issue: 4, Page: 226 - 230, Desember 30th, 2025, e-ISSN: 2502-8758
Available online at website: https://jkomtekinfo.org/ojs/index.php/komtekinfo

Combination of Support Vector Machine and Artificial Neural Network Methods in Negative Content Filtering System

M.Wira Sanjaya[1]✉, Y.Yuhandri[2] , Billy Hendrik[3]
[1,2,3]Master of Informatics Engineering, Faculty of Computer Science, Universitas Putra Indonesia YPTK, Padang, 25221, Indonesia

wirasanjaya292@gmail.com

Abstract

Local Wi-Fi network access has become a common necessity in everyday digital activities, but it is vulnerable to misuse to access negative content. This content includes pornographic material, hate speech, and violent content that can adversely affect users, especially in educational settings. For this reason, a system that is able to filter malicious content automatically and efficiently is needed. This research **aims to** design an artificial intelligence-based negative content filtering system that can be run on local network devices. The methods used include image classification using Convolutional Neural Network (CNN) and Artificial Neural Network (ANN), as well as text classification with DistilBERT and Support Vector Machine (SVM). To maintain user privacy, the model is trained using *a federated learning* approach that allows for decentralized learning. Knowledge distillation is also applied to produce lightweight models that can be run on edge devices such as routers. **The datasets** used include NSFW Image Dataset, OpenPornSet, as well as a collection of toxic comments from Reddit and Twitter. The evaluation was carried out in a simulation of a local network with 50 active devices. **The test results** showed an ANN accuracy rate of 93.4% in recognizing visual content, and SVM accuracy of 91.7% in detecting text-based hate speech. This research can be a **reference** in the application of AI-based content filtering systems for safe and responsible digital access protection.

Keyword: Negative Content, Filtering, Wifi, DPI-AI, Security network

1. Introduction

Advances in networking technology have driven increased digital connectivity in various sectors, especially through the use of local Wi-Fi networks [1]. This network is an important infrastructure in supporting online activities in the new education, government, and digital community environments [2]. However, the ease of internet access also brings serious challenges in the form of exposure to negative content, such as pornography, hate speech, violence, and disinformation [3]. This threat becomes increasingly complex when it occurs in environments with low levels of digital literacy [4]. Network security is a crucial factor in data security [5].

Currently, computer network technology is still limited to Local Area Networks [6]. Network usage continues to increase with technological advancements [7]. Various approaches have been developed to minimize these risks, one of which is through static or blacklist rule-based content filtering systems [8]. However, this method has limitations because it is not adaptive and easy for users to skip [9]. As a solution, there is a Deep Packet Inspection (DPI) and artificial intelligence (AI) based approach that is able to analyze data traffic in depth down to the payload level [10].

DPI allows the system to identify the contents of each data packet, including protocols, metadata, and actual content, making it highly effective in recognizing suspicious access patterns [11]. However, traditional DPI has its challenges in terms of compute load and limitations on encrypted traffic [12]. To overcome this, AI technology is applied to improve system intelligence through adaptive classification methods, both for visual and text content [13].

AI in this context plays an important role in recognizing anomalies, detecting explicit content, and understanding the context of user communication [14]. The use of models such as Artificial Neural Network (ANN), Convolutional Neural Network (CNN), and Support Vector Machine (SVM), combined with transformer-based Natural Language Processing (NLP) techniques such as DistilBERT, allows the system to filter content more accurately and contextually [15] [16].

Previous research has shown the effectiveness of DPI-AI hybrid systems at various scales of implementation, from schools to digital village

Submitted: 18-10-2025 | Revised: 20-11-2025 | Accepted: 18-12-2025 | Published: 30-12-2025

communities, both in technical and social aspects [17]. Other studies have also highlighted the importance of using local context-based AI [18]. The application of a filtering system on edge devices with limited resources is also an important aspect in ensuring system efficiency and sustainability [19].

Based on the explanation of the previous research, this study aims to design and evaluate a negative content filtering system that integrates AI and DPI methods in a modular architecture. The resulting system is expected to be able to work efficiently and can be run on edge devices with limited resources. The AI and DPI approach is also expected to be a practical and sustainable solution to address content security challenges on local Wi-Fi networks. The contribution of this research can be a reference in the application of AI-based content filtering systems for secure and responsible digital access protection

2. Methods

This research method consists of several main stages which are described in detail as follows. Figure 1 illustrates the research framework, outlining the key components and flow of the study.

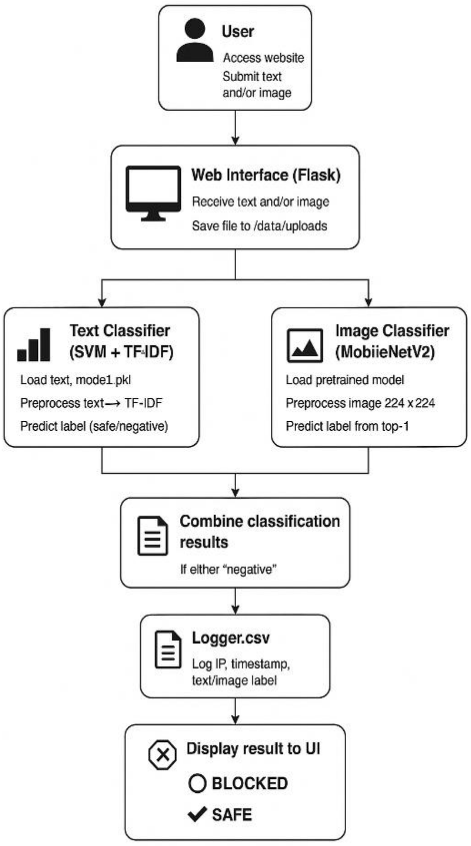

Figure 1. Research Framework

2.1 System Planning

The initial stage of the research began by designing an AI-based negative content filtering system framework that could be integrated into local Wi-Fi networks. The system is designed to be modular, efficient, and can work in real-time without disrupting network performance. The main focus of the design is to detect two types of content

2.1.1 User Interaction

The process begins when the user accesses a website built using the Flask framework [20]. On the page, users can upload text, images, or both at once [21]. The input sent can be content that will be checked for security before being allowed to be displayed or used [22].

2.1.2 Reception of Data in Web Interfaces

Flask receives data from the user, then stores it in a special directory, e.g. /data/uploads [23]. This stage ensures that files and text are securely stored on the server for later analysis [24]. At this point, the system has not yet assessed the content—only securing a copy of it for further processing [25].

2.1.3 Text Classification

If a user uploads text, the system will process it using an SVM-based Text Classifier (Support Vector Machine) that has been trained with the TF-IDF feature [26]. The process starts with preprocessing, converts it into a TF-IDF numerical representation, and then runs the SVM model to predict whether the text is safe or negative [27] [28].

2.1.4 Image Classification

If an image is uploaded, the system uses a pre-trained MobileNetV2 architecture-based Image Classifier [29]. The image was reprocessed (resized) to a size of 224×224 pixels according to the model's needs, and then predictions were made to determine the top-1 label [30]. These labels are then mapped to safe or hazardous categories [31].

2.1.5 Merger of Classification Results

Once the text and/or images have been classified, the results are combined [32]. The rule is simple: if one of the two types of content is detected as *negative*, then the final system decision is BLOCKED [33]. If both are safe, then the final status is SAFE [34].

2.1.6 Log Keeping

The system records all results to the log file logger.csv [35]. The stored data includes the user's IP address, access time (timestamp), classified labels for text and images, and final decision (BLOCKED or SAFE) [36]. This recording is useful for analysis, auditing, or evaluating system performance in the future [37].

2.1.7 Displaying Results to the Interface

Finally, the results of the decision are displayed back to the user interface. If the content is safe, it will appear in a SAFE status [38]. If it detects that it is malicious, a BLOCKED status is displayed along with a sign that access or use of the content is not allowed [39] [40].

3. Results and Discussions

3.1 Network Traffic Data

Network traffic data was obtained directly from the results of monitoring the internet activities of village public Wi-Fi network users. The focus of data collection is on the HTTP and HTTPS protocols. Figure 2 presents a sample snippet of network traffic data collected from a public Wi-Fi network.

Waktu Akses	IP Pengguna	Domain	Protokol	Ukuran Paket
2024-04-24 13:45:01	192.168.1.10	example com	HTTP	300 bytes
2024-04-24 13:46:15	192.168.1.17	memek.id	HTTPS	Diblokir
2024-04-24 13:47:30	192.168.1.21	news.com	HTTPS	Normal
2024-04-24 13:48:05	192.168.1.23	ggslot.site	HTTP	Diblokir

Figure 2 Sample Network Traffic Data Snippet from Public Wi-Fi.

Source: Data simulation by the author based on *Wireshark* and *tcpdump output structures*.

A snapshot of the data can be seen in Table 4.1, which shows a digital table of network traffic rows in real-time. Destination domain information, for example, is particularly useful in the process of classifying domain-based content that has been trained by AI models.

3.2 System Evaluation Data

After the system is successfully implemented, an evaluation process is carried out by collecting system performance data from the real environment. This data includes filtered traffic logs, network latency levels before and after filtering, the amount of content blocked, and the accuracy of the classification results calculated based on true positive, false positive, true negative, and false negative.

This image shows the results of the performance evaluation of the content classification model using the confusion matrix method as well as other evaluation metrics such as accuracy, precision, recall, F1-score, and validation accuracy using K-Fold. This evaluation is important to know how well the AI model can distinguish between safe and negative content. Figure 3 displays the results of the content classification model, highlighting its performance across different categories.

Confusion Matrix			
		Prediksi Aman	Prediksil Negatif
Real	Aman	920	30
	Negatif	40	510
Akurasi		0,95	
Presisi		0,93	
Recall		0,94	
F1-Score		0,94	
Akurasi Validasi (K-Fold)		0,94	
1		0,94	
2		0,96	
3		0,95	
4		0,94	
5		0,95	

Figure 3. Results of Content Classification Model Evaluation Using SVM and ANN

Source: Data from model testing by the author.

Table 1 illustrates the confusion matrix, detailing the model's predictions for both "Real Aman" and "Real Negative" categories, including the number of correct and incorrect classifications.

Table 1. Confusion Matrix

	Safe Predictions	Negative Predictions
Real Aman	920	30
Real Negative	40	510

Interpretation:

True Positive (TP) = 510 → Negative content predicted to be negative

True Negative (TN) = 920 → Predictably safe content

False Positive (FP) = 30 → Incorrectly classified safe content

False Negative (FN) = 40 → Negative content is misclassified as safe

Validasi K-Fold Cross Validation

The model was also evaluated using the 5-Fold Cross Validation method, with the following accuracy show in table 2:

Table 2. Validasi K-Fold Cross Validation

Fold to-	Accuracy
1	0,94
2	0,94
3	0,96
4	0,95
5	0,95

K-Fold Average Accuracy:

$$Rata - rata = \frac{4{,}74}{5} \approx 0{,}94$$

4. Conclusions

The AI model performs very well, with an overall accuracy of 95%, high recall (94%) which indicates the ability to detect negative content consistently, and a balanced F1-score (94%) which indicates predictive stability between the two classes. These results show that the constructed classification model is quite reliable and efficient in detecting negative content without sacrificing many false safe predictions. This evaluation is a strong basis for the implementation of a real-time and adaptive content filtering system in the village public network environment.

References

[1] Dwiputra, A. R., Maulana, D. A., Nurzamilah, Z., Pambudi, A. P., Nurpulaela, L., & Andromeda, S. (2025). PERAN FIBER OPTIK DALAM REVOLUSI TEKNOLOGI JARINGAN TELEKOMUNIKASI. *JATI (Jurnal Mahasiswa Teknik Infor matika)*, *9*(1), 1657-1663.

[2] Mawaddah, N., Dewi, N., Rifai, M., & Ronoatmojo, M. A. (2025). Optimalisasi Pembelajaran Daring Dalam Meningkatkan Akses Pendidikan di Daerah Terpencil. *Jurnal Kolaboratif Sains*, *8*(2), 1399-1408.

[3] IMMANUEL, I. G. D. (2024). *PERTANGGUNGJAWABAN PLATFORM DIGITAL DALAM MENGATASI KONTEN ILEGAL* (Doctoral dissertation, Universitas Islam Sultan Agung Semarang).

[4] Karengga, F. I. (2025). ANALISIS TANTANGAN PENGEMBANGAN MEDIA SERTA BAHAN AJAR BERBASIS TEKNOLOGI DALAM PENINGKATAN KOMPETENSI LITERASI DIGITAL SISWA MI. *MUBTADI: Jurnal Pendidikan Ibtidaiyah*, *6*(2), 156-169.

[5] Yanto, H., & Hadi, F. (2020). Intruder Detection Monitoring System in Computer Networks Using Snort Based Sms Alert (Sistem Monitoring Deteksi Penyusup Dalam Jaringan Komputer Menggunakan Snort Berbasis Sms Alert). Jurnal KomtekInfo, 7(2), 159-170.

[6] Wicaksana, P., Hadi, F., & Hadi, A. F. (2021). Perancangan Implementasi VPN Server Menggunakan Protokol L2TP dan IPSec Sebagai Keamanan Jaringan. Jurnal KomtekInfo, 8(3), 169-175.

[7] Ernes, R. N., & Wirawan, N. T. (2021). Pengimplementasian Artificial Inteligence pada Sistem Keamanan Locker Otomatis berbasis SMS Gateway dan Radio Frequency Identification. Jurnal KomtekInfo, 8(1), 57-65.

[8] Dawis, A. M., Rahmayanti, D., Rachman, T., Impron, A., & Kelen, Y. P. K. (2025). Pendekatan Modern Dalam Analisis Dan Desain Teknologi Informasi.

[9] Bakri, S. N., & Nasution, M. I. P. (2024). Penerapan Metodologi Rekayasa Perangkat Lunak untuk Efisiensi Pengembangan Sistem. *JSITIK: Jurnal Sistem Informasi dan Teknologi Informasi Komputer*, *3*(1), 53-66.

[10] Kushariyadi, K., Apriyanto, H., Herdiana, Y., Asy'ari, F. H., Judijanto, L., Pasrun, Y. P., & Mardikawati, B. (2024). *Artificial intelligence: Dinamika perkembangan AI beserta penerapannya*. PT. Sonpedia Publishing Indonesia.

[11] Samsumar, L. D., Nasiroh, S., Farizy, S., Anwar, C., Mursyidin, I. H., Rosdiyanto, R., ... & Prastyo, D. (2025). Keamanan Sistem Informasi: Perlindungan Data dan Privasi di Era Digital.

[12] Nirsal, N., Judijanto, L., Apriyanto, A., Susilo, A., Akbar, M. H., Sagala, L. O. H. S., ... & Paliling, A. (2025). *Riset Bidang Komputer*. PT. Sonpedia Publishing Indonesia.

[13] Sucianingtyas, R., Falistya, L. R., Pujiana, S., Prayogi, A., & Laksana, S. D. (2025). Telaah Ragam Artificial Inteligence (AI) Dalam Pendidikan. *Madani: Jurnal Ilmiah Multidisiplin*, *3*(2), 232-243.

[14] Sudirwo, S., Hadi, A., Judijanto, L., Purwandari, N., Zain, N. N. E., Rambe, K. H., ... & Yusufi, A. (2025). *Artificial Intelligence: Teori, Konsep, dan Implementasi di Berbagai Bidang*. PT. Sonpedia Publishing Indonesia.

[15] Hafidzah, P., Maryani, S., Ihsani, B. Y., Nurmiwati, N., Erwin, E., & Niswariyana, A. K. (2024, August). Penerapan Deep Learning dalam Menganalisis Sentimen di Media Sosial. In *Seminar Nasional Paedagoria* (Vol. 4, No. 1, pp. 328-339).

[16] Alifka, M. (2024). *Analisis dan Implementasi Algoritma Convolutional Neural Network pada Opinion Mining dalam Pemanfaatan Platform Chatbot di Twitter= Analysis and Implementation of Convolutional Neural Network Algorithm on Opinion Mining in the Utilization of Chatbot Platform on Twitter* (Doctoral dissertation, Universitas Hasanuddin).

[17] Azizah, S., Pt, S., Sos, M., & Commun, M. (2025). Pengembangan Masyarakat. *Pengembangan Masyarakat Berbasis Digital*, *54*.

[18] Masruroh, S. I. (2025). Kecerdasan Artifisial dan Perencanaan Pendidikan dalam Konteks Kurikulum Merdeka di Indonesia. *Jurnal Al-Kifayah: Ilmu Tarbiyah dan Keguruan*, *4*(1), 86-104.

[19] Erikasari, V. Z., Maharani, T. F., & Rilvani, E. (2025). Literature Review: Blokchain dan Edge Computing dalam Optimalisasi Sistem Terdistribusi Masa Depan. *SISFOTENIKA*, *15*(1), 43-54.

[20] Syach, U., & Edi, S. W. M. (2024). Perancangan Aplikasi Web Manajemen Data Produk Bisnis Perhiasan Berbasis Flask Dan Mongodb. *IT-Explore: Jurnal Penerapan Teknologi Informasi dan Komunikasi*, *3*(2), 162-176.

[21] Hasugian, P. S., Tuhuteru, H., Safarudin, M. S., Indahsari, A. N., Usman, A., Thaniket, R. M., ... & Hadi, A. (2025). *Pemrograman Web*. Serasi Media Teknologi.

[22] Madani, M. A. (2024). Penetration Testing untuk Menguji Sistem Keamanan pada Website. *Jeitech (Journal of Electrical Engineering And Information Technology)*, *2*(1), 33-45.

[23] Denizar, O. B. (2024). *Sistem Jaringan Monitoring Bus Operasional UKSW Berbasis Mikrokontorler* (Doctoral dissertation).

[24] Amelia, A. (2024). *Pengembangan sistem visualisasi log untuk evaluasi keamanan web server (Studi Kasus: Website Prodi Teknik Informatika UIN Syarif Hidayatullah Jakarta* (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif HIdayatullah Jakarta).

[25] Khaddafi, M., & SE, M. (2025, March). PROSES PENGEMBANGAN KONTEN MULTIMEDIA. In *Multimedia* (p. 37). Yayasan Tri Edukasi Ilmiah.

[26] Trisnawati, W., & Wibowo, A. (2024). Sentiment analysis of ICT service user using Naive Bayes classifier and SVM methods with TF-IDF text weighting. *Jurnal Teknik Informatika (JUTIF)*, *5*(3), 709-719.

[27] Bahtiar, T. (2024). *SISTEM ANALISIS SENTIMEN TERHADAP PRODUK SUNSCREEN PADA MARKETPLACE SHOPEE MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM)* (Doctoral dissertation, Universitas Islam Sultan Agung Semarang).

[28] Mola, S. A. S., Djawa, S. N. R., & Mauko, A. Y. (2025). *Text Mining: Analisis Sentimen dengan Naïve Bayes*. Kaizen Media Publishing.

[29] Muttaqin, N. H., & Widodo, A. M. (2025). Evaluation of Transfer Learning-Based Convolutional Neural Networks (InceptionV3 and MobileNetV2) for Facial Skin-Type Classification. *Jurnal Ilmu Komputer dan Informatika*, *5*(1), 11-32.

[30] Madenda, S. Komputer Vision, Kecerdasan Artifisial, dan Sistem Tertanam.

[31] Rahmatmulya, R. (2024). *Segmentasi Citra Bangunan Untuk Menentukan Tingkat Kerusakan Pasca Bencana Alam Menggunakan Convolutional Neural Network* (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).

[32] Prastyo, E. (2024). *Deteksi berita hoax dengan pendekatan Lexicon Based dan LSTM* (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).

[33] Hesaputra, A. P. (2024). *Klasifikasi Pelanggaran Undang-undang ITE menggunakan LSTM dan BILSTM* (Doctoral dissertation, Universitas Islam Indonesia).

[34] Sulianta, F. (2025). *Literasi Digital Tingkat Lanjut-Computer Security*. Feri Sulianta.

[35] Ardaisi, Y., & Yasser, R. (2025). OPTIMASI OPERASI PLTS MELALUI PEMASANGAN SOLAR TRACKER DAN MONITORING BERBASIS INTERNET: STUDI EMPIRIS PLTS PT SEG PALEMBANG. *Jurnal Desiminasi Teknologi*, 1-10.

[36] Putri, R. A. *Pemodelan algoritma random forest untuk klasifikasi log access jenis domain pada pandi (pengelola nama domain internet Indonesia)* (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif HIdayatullah Jakarta).

[37] Amelia, A. (2024). *Pengembangan sistem visualisasi log untuk evaluasi keamanan web server (Studi Kasus: Website Prodi Teknik Informatika UIN Syarif Hidayatullah Jakarta* (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif HIdayatullah Jakarta).

[38] Febriyanti, R. (2024). *Evaluasi dan Pengembangan Website GMF Safety menggunakan Metode Participatory Design* (Doctoral dissertation, Universitas Islam Indonesia).

[39] Sulianta, F. (2025). *Literasi Digital Tingkat Lanjut-Computer Security*. Feri Sulianta.

[40] Supendi, A. P. (2024). *ANALISA KERENTANAN APLIKASI WEB MENGGUNAKAN FRAMEWORK MITRE ATT&CK DENGAN METODE SIMULASI RED TEAM: STUDI KASUS DI PT. NURUL FIKRI CIPTA INOVASI* (Doctoral dissertation, Sekolah Tinggi Teknologi Terpadu Nurul Fikri).

Biographies of Authors

	**M. Wira Sanjaya** is a postgraduate student at Putra Indonesia YPTK University Padang, working as a lecturer and researcher. He was born on May 17, 1999, in Kerinci. He is majoring in computer science and pursuing a master's degree at Putra Indonesia YPTK University Padang. He can be contacted via email at wirasanjaya292@gmail.com. He lives in Ujung Pasir village, Tanah Cogok sub-district, Kerinci district, Jambi province.
	**Yuhandri** was born in Tanjung Alam on May 15. He is an Assistant Professor in Faculty of Computer Science, Universitas Putra Indonesia YPTK. He received the Bachelor Degree in Informatics Management and Master Degree in Information Tecnology in 1992 and 2006 from Universitas Putra Indonesia YPTK. Moreover, he completed his Doctorate of Information Technology as Informatics Medical Image expertise from Gunadarma University in April 2017. He is a lecturer at the Faculty of Computer Science, Universitas Putra Indonesia YPTK. Scopus Id is 57193430920. E-mail: yuyu@upiyptk.ac.id
	**Billy Hendrik** born on March 18, 1983. Bachelor's degree in the Faculty of Computer Science, Computer Engineering Study Program, then completed his Master's degree in Information Technology Study Program at Universitas Putra Indonesia YPTK Padang and Doctoral Education at Universiti Kebangsaan Malaysia. He teaches at Universitas Putra Indonesia "YPTK" Padang. He has knowledge in the fields of Technology, Education, Robotics. He is currently active as Deputy Head of the Student Affairs and Alumni Bureau. Chairman of the Madinah Islamic Education Foundation. The author actively conducts research in the field of Robotics and its Application in the field of Education. The author also actively writes articles in various scientific journals both nationally and internationally. Phone: 085272557211 Email: billy_hendrik@upiyptk.ac.id