

Analysis of the Impact of Violent Content on Social Media on Adolescent Cyberpsychology Using Support Vector Machine and Random Forest

Wulandari Febriani ^{1*}, Mambang ^{2*}, Septyan Eka Prastya ^{3*}, Billy Sabella ^{4**}, Finki Dona Marleny ^{3***}

^{*} Teknologi Informasi, Universitas Sari Mulia

^{**} Teknologi Informasi, Politeknik Negeri Tanah Laut

^{***} Teknik Informatika, Universitas Muhammadiyah Banjarmasin

wulandari23febriani@gmail.com ¹,

Article Info

Article history:

Received 2025-10-02

Revised 2026-01-11

Accepted 2026-01-30

Keywords:

*Cyberpsychology,
Violent Content,
Multi-label Classification,
Support Vector Machine,
Random Forest.*

ABSTRACT

Adolescent exposure to violent content on social media has emerged as a critical issue due to its potential impact on mental health and cyberpsychological well-being. This study aims to classify multiple cyberpsychological impacts experienced by adolescents as a result of exposure to violent content on social media using a multi-label machine learning approach. A quantitative method was employed using self-reported data collected from 550 Indonesian adolescents aged 12–18 years through an online questionnaire. Psychological impacts were measured using adapted instruments from the Depression Anxiety Stress Scales (DASS-21) and cyberpsychology scales, then transformed into multi-label targets. Support Vector Machine (SVM) and Random Forest algorithms were implemented using a One-vs-Rest strategy. Model performance was evaluated using Hamming Loss, precision, recall, and Macro F1-score. The results indicate that SVM outperformed Random Forest with a Hamming Loss of 23.16% and a Macro F1-score of 0.42, particularly in predicting dominant labels such as anxiety and decreased self-confidence. However, both models showed limited performance in predicting minority labels such as depression and academic decline due to data imbalance. These findings highlight the importance of handling imbalanced data in cyberpsychology-based machine learning research and demonstrate the potential of multi-label classification in representing the complexity of psychological impacts of digital violence on adolescents.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Perkembangan teknologi digital telah meningkatkan intensitas penggunaan media sosial di kalangan remaja. Laporan We Are Social dan Hootsuite (2024) mencatat lebih dari 4,8 miliar pengguna aktif media sosial secara global, dengan remaja sebagai kelompok pengguna paling intensif [1]. Di Indonesia, sekitar 84% remaja memiliki akses internet rutin dan menghabiskan lebih dari tiga jam per hari di platform seperti Instagram, TikTok, dan YouTube [2].

Di samping manfaat sosial dan edukatif, paparan konten negatif, terutama konten kekerasan yang menjadi ancaman serius bagi kesejahteraan psikologis remaja [3]. Konten seperti cyberbullying, hate speech, body shaming, dan pelecehan seksual telah dikaitkan dengan berbagai gangguan

psikologis, termasuk kecemasan, stres, gangguan tidur, penurunan prestasi akademik, dan depresi [4][5]. Penelitian sebelumnya menunjukkan bahwa paparan konten kekerasan digital berpengaruh signifikan terhadap kondisi emosional dan psikologis remaja [6].

Sejumlah penelitian telah menerapkan pendekatan machine learning dalam analisis konten digital, namun umumnya masih terbatas pada klasifikasi biner atau hanya memodelkan satu jenis dampak psikologis [7][8]. Penelitian Perera dan Fernando [6] menggunakan supervised machine learning untuk deteksi cyberbullying, namun hanya fokus pada klasifikasi biner (bullying vs non-bullying) tanpa mengeksplorasi dampak psikologis yang ditimbulkan. Sementara itu, Yi dan Zubiaga [5] mengembangkan sistem

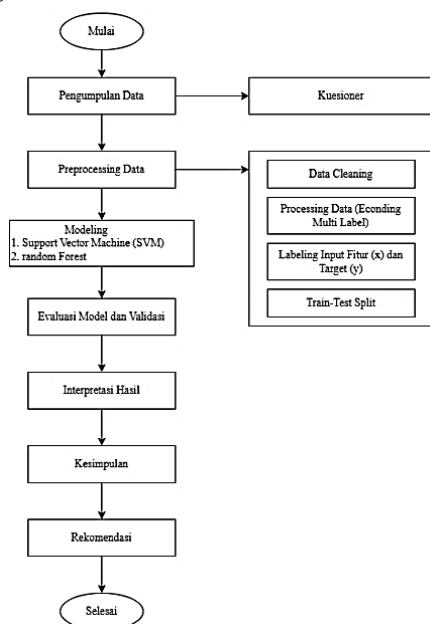
deteksi cyberbullying berbasis sesi, tetapi tidak memodelkan multipel dampak psikologis secara simultan. Padahal, dampak psikologis akibat paparan konten kekerasan pada remaja sering bersifat multipel dan terjadi secara simultan [9]. Pendekatan klasifikasi multi-label memungkinkan satu sampel diklasifikasikan ke dalam beberapa kategori dampak psikologis sekaligus, sehingga lebih merepresentasikan kompleksitas kondisi tersebut [10]. Hingga saat ini, belum ada penelitian yang secara khusus mengintegrasikan perspektif cyberpsychology dengan multi-label classification untuk memodelkan dampak psikologis multipel pada remaja Indonesia akibat paparan konten kekerasan di media sosial.

Kebaruan penelitian ini terletak pada penerapan pendekatan multi-label classification untuk memodelkan dampak cyberpsychological remaja secara simultan. Secara khusus, penelitian ini bertujuan untuk: (1) menganalisis distribusi berbagai dampak cyberpsychological pada remaja, (2) mengevaluasi dan membandingkan kinerja algoritma SVM dan Random Forest dalam tugas klasifikasi multi-label, serta (3) mengkaji pengaruh ketidakseimbangan label terhadap performa model.

II. METODE

A. Desain dan Alur Penelitian

Alur metodologi penelitian ditunjukkan pada Gambar 1, yang mencakup tahapan pengumpulan data, *preprocessing*, pemodelan, dan evaluasi.



Gambar 1. Alur Penelitian,
Sumber: Data kuesioner, 2025

B. Pengumpulan Data

Data dikumpulkan melalui kuesioner daring menggunakan Google Form yang disebarakan kepada remaja berusia 12–18 tahun di Indonesia. Kuesioner terdiri atas tiga bagian utama: (1) data demografis, (2) pola penggunaan media sosial dan paparan konten kekerasan, serta (3) dampak psikologis yang dirasakan responden.

Pengukuran dampak psikologis dilakukan dengan mengadaptasi Depression Anxiety Stress Scales (DASS-21) [12] dan Cyberbullying Impact Scale yang telah digunakan dalam penelitian sebelumnya. Instrumen DASS-21 terdiri dari 21 item yang mengukur tiga dimensi (depression, anxiety, stress) dengan skala Likert 4 poin (0 = tidak pernah, 3 = sangat sering). Cyberbullying Impact Scale mengukur dampak seperti penurunan kepercayaan diri, gangguan tidur, masalah hubungan sosial, dan penurunan prestasi akademik menggunakan skala yang sama.

Skor pada setiap dimensi dihitung dengan menjumlahkan nilai item, kemudian dikalikan 2 sesuai pedoman DASS-21. Kategori dampak ditentukan berdasarkan cut-off score berikut: untuk anxiety (normal: 0-7, ringan: 8-9, sedang: 10-14, berat: 15-19, sangat berat: ≥ 20); untuk depression (normal: 0-9, ringan: 10-13, sedang: 14-20, berat: 21-27, sangat berat: ≥ 28); untuk stress (normal: 0-14, ringan: 15-18, sedang: 19-25, berat: 26-33, sangat berat: ≥ 34). Untuk dimensi cyberpsychology lainnya, cut-off score disesuaikan berdasarkan distribusi kuartil data.

Responden dikategorikan mengalami suatu dampak psikologis apabila skor berada pada kategori sedang atau lebih tinggi. Setiap kategori dampak kemudian dipetakan sebagai label biner (1 = terdampak jika skor $\geq$ sedang, 0 = tidak terdampak jika skor $<$ sedang), sehingga memungkinkan terbentuknya lebih dari satu label untuk setiap responden.

Uji reliabilitas menghasilkan nilai Cronbach's Alpha sebagai berikut: anxiety ($\alpha = 0,82$), depression ($\alpha = 0,79$), stress ($\alpha = 0,78$), penurunan kepercayaan diri ($\alpha = 0,81$), gangguan tidur ($\alpha = 0,76$), masalah hubungan sosial ($\alpha = 0,74$), dan penurunan prestasi akademik ($\alpha = 0,73$). Semua nilai berada di atas threshold 0,70, yang menunjukkan konsistensi internal yang memadai untuk penelitian eksploratori.

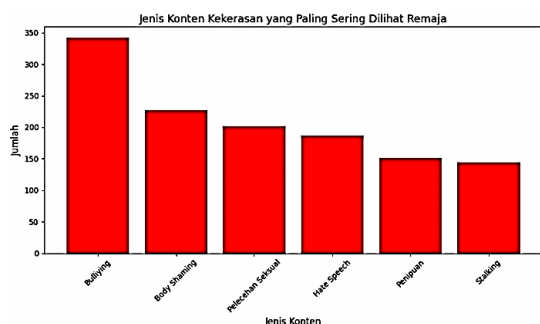
C. Karakteristik Sampel

Dari 550 respons awal, sebanyak 485 data dinyatakan valid setelah proses pembersihan data. Karakteristik demografis responden disajikan pada Tabel 1.

TABEL 1.
KARAKTERISTIK DEMOGRAFIS RESPONDEN (N=550) SUMBER: PENELITI, 2025

Variable	Kategori	Jumlah(n)	Presentase(%)
Jenis Kelamin	Perempuan	310	54,6
	Laki-laki	258	45,5
Usia	12-14 tahun	126	22,1
	15-16 tahun	188	33,1
	17-18 tahun	254	44,7
Durasi Media Sosial	< 1 jam	60	10,6
	1-3 jam	197	34,7
	>3 jam	331	58,3

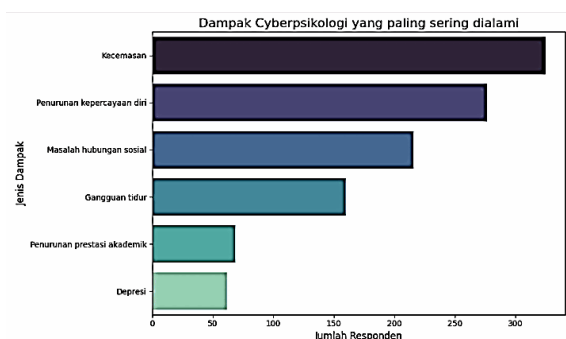
Distribusi jenis konten kekerasan dan dampak psikologis divisualisasikan pada Gambar 2.



Gambar 2. Distribusi Jenis Konten Kekerasan
Sumber: Data kuesioner, 2025

D. Preprocessing

Fitur kategorikal seperti jenis konten kekerasan dan platform media sosial dikodekan menggunakan One-Hot Encoding. Fitur numerik yang digunakan dalam model meliputi: (1) skor total setiap dimensi psikologis dari DASS-21 dan Cyberbullying Impact Scale (7 fitur), (2) durasi penggunaan media sosial per hari dalam menit (1 fitur), (3) frekuensi paparan konten kekerasan per minggu (1 fitur), dan (4) usia responden (1 fitur), sehingga total terdapat 10 fitur numerik. Fitur numerik berupa skor psikologis dinormalisasi menggunakan Min-Max Scaler ke dalam rentang [0,1] untuk menghindari bias akibat perbedaan skala antar fitur. Setelah One-Hot Encoding pada fitur kategorikal (jenis konten: 5 kategori, platform: 4 kategori), total dimensi fitur mencapai 19 fitur



Variabel target dampak psikologis yang bersifat multi-label ditransformasikan menggunakan MultiLabelBinarizer. Data kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan stratified splitting berdasarkan kombinasi label untuk menjaga proporsi distribusi label yang tidak seimbang

E. Pemodelan dan Evaluasi

Pemodelan dilakukan menggunakan algoritma SVM (kernel linear) dan Random Forest (100 trees) dengan pendekatan One-vs-Rest untuk menangani multi-label classification. Strategi One-vs-Rest dipilih karena: (1) jumlah label yang relatif terbatas (6 label) memungkinkan komputasi yang efisien, (2) setiap label memiliki karakteristik yang

dapat diperlakukan secara independen, (3) pendekatan ini memberikan fleksibilitas dalam mengoptimalkan parameter untuk setiap label secara terpisah, dan (4) interpretabilitas model lebih baik dibandingkan strategi lain seperti Binary Relevance atau Classifier Chains [14].

Sebagai pembandingan, digunakan model baseline DummyClassifier dengan strategi most_frequent yang memprediksi label mayoritas untuk setiap kelas secara terpisah. Evaluasi performa dilakukan menggunakan *label-based accuracy*, *Hamming Loss*, serta *macro-averaged precision*, *recall*, dan *F1-score* untuk memberikan gambaran kinerja yang seimbang antar label.

III. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini dimulai dari pengumpulan data yang dilakukan melalui penyebaran kuesioner secara daring menggunakan *platform* seperti *Google Form*. Sasaran utama responden adalah remaja berusia 12 hingga 18 tahun. Berikut data demografinya. Berikut adalah Tabel hasil Kuesioner berdasarkan data demografi responden.

A. Distribusi Data dan Ketidakseimbangan Label

Analisis awal menunjukkan ketidakseimbangan ekstrem pada label dampak psikologis. Seperti ditunjukkan pada Tabel 2, kecemasan merupakan kelas mayoritas (58,2%), diikuti penurunan kepercayaan diri (51,8%). Sementara itu, depresi dan penurunan prestasi akademik merupakan kelas minoritas (<10%). Distribusi ini menjadi faktor kritis yang memengaruhi performa model.

TABEL 2.
DISTRIBUSI LABEL DAMPAK PSIKOLOGIS SUMBER : PENELITIAN, 2025

Dampak Psikologis	Jumlah Sampel	Persentase (%)
Kecemasan	282	58,2
Penurunan Kepercayaan Diri	251	51,8
Gangguan Tidur	194	40,0
Masalah Hubungan Sosial	163	33,6
Depresi	41	8,5
Penurunan Prestasi Akademik	38	7,8

Berdasarkan Tabel 2, algoritma Support Vector Machine menunjukkan kinerja terbaik dengan nilai Hamming Loss terendah dan Macro F1-score tertinggi dibandingkan model lainnya, sehingga lebih efektif dalam menangani klasifikasi multi-label pada data tidak seimbang.

B. Performa Model Klasifikasi

Kinerja model dievaluasi menggunakan metrik Hamming Loss dan Macro F1-score untuk menilai kemampuan klasifikasi secara menyeluruh pada data multi-label pada tabel 3.

TABEL 3.
PERFORMA MODEL KLASIFIKASI MULTI-LABEL SUMBER: PENELITI, 2025

Model	Akurasi (%)	Hamming Loss (%)	Macro F1-Score
Baseline (Dummy)	10,12	31,45	0,09
SVM	17,89	23,16	0,42
Random Forest	14,73	27,01	0,38

Hasil pada Tabel 3 menunjukkan bahwa kedua algoritma memiliki performa yang baik pada label dominan, namun mengalami penurunan kinerja pada label minoritas. Nilai akurasi label-based yang tampak rendah (SVM: 17,89%, RF: 14,73%) perlu dipahami dalam konteks multi-label classification. Akurasi label-based mengukur persentase sampel yang semua labelnya diprediksi dengan benar secara sempurna, sehingga metrik ini sangat ketat dan cenderung rendah pada kasus multi-label dengan banyak label [14]. Metrik yang lebih relevan untuk evaluasi adalah Hamming Loss dan Macro F1-score. Hamming Loss SVM sebesar 23,16% menunjukkan bahwa rata-rata kesalahan prediksi per label hanya 23,16%, yang mengindikasikan performa yang cukup baik mengingat kompleksitas data dengan 6 label dan distribusi yang tidak seimbang.

Hal ini mengindikasikan adanya bias model terhadap kelas mayoritas, yang merupakan karakteristik umum pada data dengan distribusi tidak seimbang [13]. Model SVM menunjukkan Macro F1-score sebesar 0,42, yang lebih tinggi 16,7% dibandingkan baseline (0,09) dan 10,5% dibandingkan Random Forest (0,38), menunjukkan bahwa SVM mampu menangani kompleksitas multi-label dengan lebih baik.

C. Analisis Performa per Label

Analisis lebih lanjut dilakukan untuk mengevaluasi kinerja masing-masing model pada setiap label dampak psikologis.

TABEL 4.
PERFORMA F1-SCORE PER LABEL SUMBER: PENELITI, 2025

Dampak Psikologis	SVM	Random Forest
Kecemasan	0,80	0,78
Penurunan Kepercayaan Diri	0,84	0,77
Gangguan Tidur	0,61	0,58
Masalah Sosial	0,55	0,53
Depresi	0,22	0,25
Penurunan Prestasi Akademik	0,19	0,21

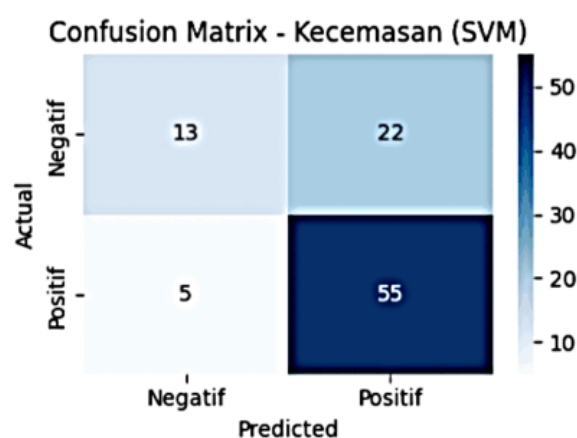
Hasil pada Tabel 4 menunjukkan disparitas performa yang signifikan antar label. SVM mencapai F1-score tertinggi pada label penurunan kepercayaan diri (0,84) dan kecemasan (0,80), yang merupakan kelas mayoritas dengan prevalensi

>50%. Sebaliknya, performa pada label minoritas sangat rendah: depresi (0,22) dan penurunan prestasi akademik (0,19). Fenomena ini terjadi karena algoritma learning cenderung mengoptimalkan fungsi loss secara global, yang didominasi oleh kontribusi kelas mayoritas [13]. Dengan hanya 41 sampel depresi (8,5%) dan 38 sampel penurunan prestasi akademik (7,8%) dari total 485 data, model tidak memiliki cukup contoh untuk mempelajari pola yang robust pada kelas minoritas.

Rendahnya performa pada label depresi dapat dijelaskan oleh sifat gejala depresi yang cenderung laten dan kompleks, sehingga tidak selalu muncul secara langsung sebagai respons terhadap paparan konten kekerasan. Depresi sering memerlukan waktu inkubasi yang lebih lama dan dipengaruhi oleh faktor-faktor lain seperti genetik, lingkungan keluarga, dan kondisi sosial-ekonomi yang tidak terukur dalam penelitian ini. Temuan ini sejalan dengan teori psikologi perkembangan yang menyatakan bahwa manifestasi depresi pada remaja bersifat multidimensional dan sulit terdeteksi melalui pengukuran tunggal.

Fenomena ini mencerminkan bias algoritma terhadap kelas mayoritas sebagaimana dijelaskan dalam teori learning from imbalanced data, di mana model cenderung mengoptimalkan performa global dengan mengorbankan sensitivitas terhadap kelas minoritas.

Gambar 4 menunjukkan *Confusion Matrix* untuk prediksi label kecemasan oleh model SVM, mengungkap jumlah *true positive*, *false positive*, *true negatif*, dan *false negatif*.



Gambar 4. *Confusion Matrix* SVM
Sumber : Penulis 2025

F. Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan yang perlu dipertimbangkan dalam interpretasi hasil:

- 1) Data self-reported: Penggunaan kuesioner self-report berpotensi mengandung bias seperti social desirability bias dan recall bias, yang dapat mempengaruhi akurasi pengukuran dampak psikologis.
- 2) Ketidakseimbangan label ekstrem: Disparitas jumlah sampel antar label (dari 8,5% hingga 58,2%)

menyebabkan bias model terhadap kelas mayoritas dan menurunkan sensitivitas terhadap kelas minoritas.

- 3) Terbatasnya teknik penanganan imbalance: Penelitian ini hanya menggunakan dua algoritma tanpa menerapkan teknik resampling seperti SMOTE, ADASYN, atau cost-sensitive learning yang dapat meningkatkan performa pada kelas minoritas.
- 4) Validasi cross-cultural: Instrumen DASS-21 dan Cyberbullying Impact Scale diadaptasi dari konteks Western, sehingga mungkin tidak sepenuhnya menangkap nuansa pengalaman cyberpsychological remaja Indonesia.
- 5) Keterbatasan fitur: Model tidak memasukkan variabel moderator seperti resiliensi psikologis, dukungan sosial, atau karakteristik kepribadian yang dapat memengaruhi dampak paparan konten kekerasan.

Untuk penelitian mendatang, disarankan untuk: (1) menggunakan teknik resampling seperti SMOTE atau ensemble methods untuk menangani data tidak seimbang, (2) mengeksplorasi algoritmi lain seperti XGBoost, deep learning, atau hybrid approaches, (3) melakukan validasi instrumen secara cross-cultural, (4) mengintegrasikan variabel moderator dan mediator dalam model, dan (5) mempertimbangkan pendekatan longitudinal untuk menangkap dinamika temporal dampak cyberpsychological.

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa paparan konten kekerasan di media sosial berkorelasi dengan dampak cyberpsychological yang bersifat multipel pada remaja. Pendekatan *multi-label classification* mampu merepresentasikan kompleksitas tersebut, namun performa model sangat dipengaruhi oleh ketidakseimbangan distribusi label. Algoritma SVM menunjukkan performa yang lebih baik dibandingkan Random Forest dalam hal kesalahan prediksi label, meskipun keduanya mengalami kegagalan pada kelas minoritas. Temuan ini menegaskan pentingnya strategi penanganan data tidak seimbang dalam penelitian cyberpsychology berbasis machine learning.

ETIKA PENELITIAN

Penelitian ini telah memperoleh persetujuan etik dari institusi terkait. Seluruh responden berpartisipasi secara sukarela dengan persetujuan sadar (*informed consent*).

DAFTAR PUSTAKA

- [1] S. Mader, D. Costantini, A. Fahr, and M. Delgrande, "Social Science & Medicine The effect of social media use on adolescents' subjective well-being : Longitudinal evidence from Switzerland," *Soc. Sci. Med.*, vol. 365, no. December 2024, p. 117595, 2025, doi: 10.1016/j.socscimed.2024.117595.
- [2] M. Nur Cahya, W. Ningsih, and A. Lestari, "Dampak Media Sosial terhadap Kesejahteraan Psikologis Remaja: Tinjauan Pengaruh Penggunaan Media Sosial pada Kecemasan dan Depresi Remaja," *J. Sos. Teknol.*, vol. 3, no. 8, pp. 704–706, 2023, doi: 10.59188/jurnalsostech.v3i8.917.
- [3] S. W. Fajriani *et al.*, "Cyberspace: Dampak Penyimpangan Perilaku Komunikasi Remaja Cyberspace: The Impact of Adolescent Communication Behavior Deviation," *J. Ilmu Pengetah. dan Teknol. Komun.*, vol. 23, no. 1, pp. 63–78, 2021, [Online]. Available: <http://dx.doi.org/10.33169/iptekom.23.1.2021.63-78>
- [4] L. Winstone, S. Jamal, and B. Mars, "Cyberbullying Perpetration and Victimization as Risk Factors for Self-Harm: Results From a Longitudinal Cohort Study of 13–14-Year-Olds in England," *J. Adolesc. Heal.*, vol. 75, no. 2, pp. 298–304, 2024, doi: 10.1016/j.jadohealth.2024.04.004.
- [5] P. Yi and A. Zubiaga, "Session-based cyberbullying detection in social media: A survey," *Online Soc. Networks Media*, vol. 36, no. August 2022, p. 100250, 2023, doi: 10.1016/j.osnem.2023.100250.
- [6] A. Perera and P. Fernando, "Cyberbullying Detection System on Social Media Using Supervised Machine Learning," *Procedia Comput. Sci.*, vol. 239, no. 2023, pp. 506–516, 2024, doi: 10.1016/j.procs.2024.06.200.
- [7] I. Nissar *et al.*, "An Intelligent Healthcare System for Automated Diabetes Diagnosis and Prediction using Machine Learning," *Procedia Comput. Sci.*, vol. 235, pp. 2476–2485, 2024, doi: 10.1016/j.procs.2024.04.233.
- [8] K. C. Onyelowe, C. B. Mahesh, B. Srikanth, C. Nwa-david, J. Obimba-wogu, and J. Shakeri, "Support vector machine (SVM) prediction of coefficients of curvature and uniformity of hybrid cement modified unsaturated soil with NQF inclusion," *Clean. Eng. Technol.*, vol. 5, p. 100290, 2021, doi: 10.1016/j.clet.2021.100290.
- [9] S. R. Khawaja *et al.*, "Journal of Hand Surgery Global Online The Impact of Social Media for Hand Surgeons : A Prevalence and Correlation Study With Online and Academic Reputations," *J. Hand Surg. Glob. Online*, pp. 1–6, 2024, doi: 10.1016/j.jhsg.2024.11.001.
- [10] K. Oka, J. He, Y. Honda, and Y. Hijioka, "Random forest analysis of the relative importance of meteorological indicators for heatstroke cases in Japan based on the degree of severity and place of occurrence," *Environ. Res.*, vol. 263, no. P2, p. 120066, 2024, doi: 10.1016/j.envres.2024.120066.
- [11] G. Obaido *et al.*, "Machine Learning with Applications Supervised machine learning in drug discovery and development : Algorithms , applications , challenges , and prospects," *Mach. Learn. with Appl.*, vol. 17, no. July, p. 100576, 2024, doi: 10.1016/j.mlwa.2024.100576.
- [12] Lovibond, S. H., & Lovibond, P. F. (1995). Manual for the Depression Anxiety Stress Scales. Psychology Foundation.
- [13] He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering.
- [14] Zhang, M. L., & Zhou, Z. H. (2014). A review on multi-label learning algorithms. IEEE Transactions on Knowledge and Data Engineering.
- [15] Kuss, D. J., & Griffiths, M. D. (2017). Social networking sites and addiction. International Journal of Environmental Research and Public Health.
- [16] APIII. (2023). Survey Pengguna Internet di Indonesia 2022-2023.