



OPTIMALISASI LAYANAN HOTLINE KESEHATAN MELALUI SISTEM INFORMASI BERBASIS TEKNOLOGI PEMROSESAN BAHASA ALAMI

Muhammad Imam Dinata^{1*}, Muhammad Rizkillah²

^{1,2}Sistem dan Teknologi Informasi, Universitas Muhammadiyah Mataram

email: imamdinata@ummatt.ac.id^{1*}

Abstrak: Keterbatasan tenaga profesional dan subjektivitas penilaian membuat layanan *hotline* kesehatan mental di Indonesia rentan terhadap bias dan keterlambatan dalam menangani kasus krisis. Penelitian ini mengembangkan sistem klasifikasi tingkat urgensi percakapan berbasis *Natural Language Processing* (NLP) menggunakan model BERT, yang diintegrasikan ke dalam aplikasi web *real-time* dengan arsitektur Python-Flask dan database MySQL. Dataset terdiri dari 150 percakapan yang dilabeli secara manual dalam tiga kategori urgensi: rendah, sedang, dan tinggi. Model dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil menunjukkan akurasi 70%, dengan F1-score 1.00 (rendah), 0.47 (sedang), dan 0.61 (tinggi). Sistem terbukti mampu mempercepat proses identifikasi kasus dan mendapat respons positif dari operator *hotline*. Kontribusi utama penelitian ini terletak pada pengembangan sistem deteksi urgensi berbasis bahasa Indonesia yang teruji fungsional dan siap diimplementasikan dalam konteks layanan psikososial digital di negara berkembang.

Kata Kunci : Hotline, Mental Health Services, Klasifikasi, BERT, NLP.

PENDAHULUAN

Gangguan kesehatan mental menjadi masalah global yang mendesak, terutama di negara berkembang (*Low and Middle Income Countries*) yang memiliki keterbatasan akses layanan psikologis. WHO mencatat lebih dari 264 juta orang mengalami depresi, dan lebih dari 800.000 kasus bunuh diri setiap tahun [4]. Untuk menjawab kondisi ini, *hotline* kesehatan mental hadir sebagai layanan cepat, anonim, dan krusial dalam menghubungkan individu dengan bantuan profesional, khususnya bagi populasi rentan seperti penderita gangguan jiwa dan mereka yang tinggal di daerah terpencil [4].

Meski penting, efektivitas *hotline* masih bergantung pada penilaian subjektif staf terhadap urgensi percakapan, yang rentan terhadap bias dan keterlambatan [7], [19]. Teknologi *Natural Language Processing* (NLP) menawarkan solusi dengan mendeteksi sinyal krisis secara otomatis dari isi percakapan. NLP mampu mengidentifikasi pola emosional seperti kecemasan, depresi, dan niat bunuh diri [10], serta dapat diintegrasikan dengan *Electronic Health Record* (EHR) untuk memperkuat keputusan klinis [13]. Di Indonesia, NLP berpotensi besar untuk menjangkau daerah dengan keterbatasan layanan. Selain mendeteksi emosi tersembunyi [9] teknologi ini dapat mengotomatisasi notifikasi tingkat urgensi, mempercepat intervensi [2]. Namun, tantangan teknis seperti infrastruktur terbatas, kurangnya SDM ahli, dan keragaman budaya lokal masih menjadi hambatan [8], [12], [14]. Penelitian ini mengembangkan sistem deteksi urgensi percakapan pada layanan *hotline* kesehatan mental berbasis NLP menggunakan model BERT (*Bidirectional Encoder Representations from Transformers*). Tujuannya adalah mengembangkan aplikasi *real-time* untuk klasifikasi urgensi, menguji efektivitas sistem dalam mendeteksi krisis psikologis, mengidentifikasi tantangan implementasi, serta mengevaluasi kontribusinya terhadap peningkatan respons layanan.

TINJAUAN PUSTAKA

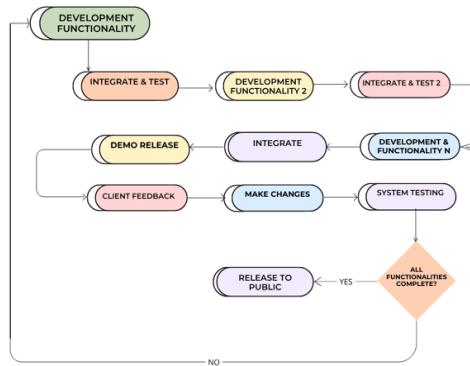
Layanan hotline kesehatan mental telah terbukti efektif dalam menangani krisis psikologis, khususnya di negara berkembang yang menghadapi keterbatasan sumber daya [1]. Di Indonesia, tekanan demografis memperkuat urgensi peningkatan layanan bagi remaja melalui pendekatan digital [17]. Salah satu pendekatan yang menjanjikan adalah pemanfaatan teknologi NLP, yang dinilai mampu meningkatkan akses dan mendeteksi kondisi psikologis secara cepat [3]. NLP telah digunakan secara luas untuk menganalisis percakapan berbasis teks, terutama dalam mengenali sinyal emosional seperti kecemasan dan keputusasaan secara real-time [16]. Teknologi ini unggul dalam mendeteksi isyarat halus yang kerap terlewat oleh penilaian manusia [8], dan mendukung proses pengambilan keputusan pada layanan psikologis berbasis teks [9]. Model NLP lanjutan seperti BERT mampu meningkatkan ketepatan klasifikasi tekanan psikologis [3], dengan performa yang baik dalam memahami ekspresi linguistik kompleks dalam data klinis [17]. Dalam penerapannya, BERT menunjukkan akurasi tinggi dalam membedakan informasi relevan dalam percakapan krisis [11], serta memperkuat kapasitas deteksi terhadap tekanan emosional akut [18]. Namun, tantangan etis seperti potensi bias algoritmik dan keterbatasan data pelatihan masih menjadi isu penting [15]. Beberapa studi menekankan perlunya kerangka etis dan adaptasi lokal, termasuk melalui prompt engineering berbasis konteks budaya [1], [18]. Validasi lokal sangat penting untuk menjaga keandalan sistem NLP dalam konteks nyata. Tanpa data lokal yang memadai, sistem dapat mengalami kesalahan klasifikasi dan gagal memenuhi kebutuhan pengguna [5]. Oleh karena itu, penguatan model berbasis data lokal menjadi strategi utama untuk memastikan akurasi dan relevansi deteksi krisis [1].



METODE

Pendekatan Penelitian

Penelitian ini menerapkan pendekatan *Research and Development* (R&D) dengan model pengembangan *Agile* dapat dilihat pada gambar 1 dibawah ini, metode pengembangan ini mendukung proses iteratif dan adaptif dalam pembangunan sistem berbasis umpan balik pengguna. Jenis penelitian ini bersifat terapan, dengan fokus pada pengembangan *prototipe* sistem cerdas yang mampu mengklasifikasikan tingkat urgensi percakapan berbasis teks dalam layanan *hotline* kesehatan mental. Pendekatan ini memungkinkan integrasi antara aspek teknis dan kebutuhan nyata di lapangan.



Gambar 1. Model pengembangan agile

Sumber: USF Health

Analisis Kebutuhan

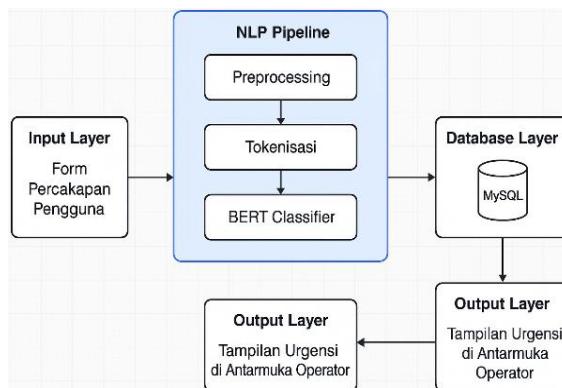
Analisis kebutuhan sistem dilakukan melalui studi literatur, wawancara semi-terstruktur dengan operator layanan *hotline*, serta penelaahan terhadap *Urgency Numeric Rating Scale* (NRS) sebagai acuan pelabelan keurgensian. Selanjutnya setelah pelabelan dilakukan maka tahap selanjutnya yaitu proses data tersebut dengan metode pemrosesan bahasa alami (NLP) sehingga hasil yang dihasilkan nantinya diharapkan akan mampu melakukan klasifikasi urgensi secara *real-time*, meminimalkan subjektivitas penilaian manual, serta menyediakan antarmuka yang intuitif dan dapat dioperasikan dengan cepat dalam kondisi darurat.

Dataset dan Pra-pemrosesan

Dataset yang digunakan terdiri dari 150 percakapan teks layanan *hotline*, yang telah dilabeli oleh pakar ke dalam tiga kategori urgensi: rendah, sedang, dan tinggi. Proses pelabelan dilakukan secara manual dengan dibantu oleh psikolog dan operator berpengalaman untuk memastikan bushwa pelabelan sudah baik dan benar. Selanjutnya data kemudian melalui proses pra-pemrosesan standar, yang mencakup *tokenisasi* untuk memisahkan kalimat yang disampaikan dalam bentuk *teks* menjadi unit kata sendiri, *stemming* untuk menghilangkan kata yang berimbuhan dan menyederhanakan bentuk kata, penghapusan kata umum dan kata hubung agar menjadi kata dasar (*stopword removal*), dan normalisasi teks untuk menyamakan format kalimat menjadi kata baku. Proses ini dilaksanakan menggunakan *library* Python berbasis Natural Language Processing (NLP) seperti NLTK dan *tokenizer* dari model BERT. Tahapan ini bertujuan untuk menghasilkan data masukan agar bersih, menghindari *noise* dan konsisten sebagai masukan ke dalam pembuatan model klasifikasi.

Arsitektur Sistem dan Pengembangan Model

Model klasifikasi utama yang digunakan adalah BERT, yang dilatih awal (*pre-trained*) dan kemudian disesuaikan (*fine-tuned*) pada data lokal dengan skenario klasifikasi *multi-class*. Proses klasifikasi dimulai dari transformasi input teks melalui tokenizer BERT, dilanjutkan dengan ekstraksi *embedding* kontekstual melalui *encoder*, dan diakhiri dengan klasifikasi tingkat urgensi menggunakan lapisan *dense* dengan aktivasi *softmax*. Arsitektur sistem terdiri atas *backend* berbasis Python dengan *framework* Flask untuk menyajikan hasil prediksi secara *real-time*, sistem basis data MySQL untuk menyimpan hasil klasifikasi, dan antarmuka web yang dirancang agar mudah diakses oleh operator *hotline*. Alur sistem secara keseluruhan digambarkan pada gambar 2. Diagram alur sistem klasifikasi tingkat urgensi berbasis BERT dibawah ini.



Gambar 2. Diagram alir sistem klasifikasi tingkat urgensi berbasis BERT

Evaluasi Model dan Sistem

Evaluasi model yang dilakukan pada penitian ini dalam dua tingkat: kinerja model klasifikasi dan fungsionalitas sistem secara keseluruhan. Model dievaluasi menggunakan pembagian data *train-test* dengan proporsi 80:20 berdasarkan teknik *stratified sampling* untuk menjaga keseimbangan distribusi label. Pengukuran model yang digunakan pada penitian ini mencakup *Accuracy*, *precision*, *recall*, dan *F1-score*.

a. Accuracy

Accuracy menjelaskan mengenai keseluruhan model yang telah dibuat dan selanjutnya akan memprediksi nilai yang benar, Rumus yang digunakan untuk perhitungan *Accuracy*, yaitu [10]:

$$\text{Accuracy} = \frac{\text{Total Prediksi benar}}{\text{Total data uji}} \times 100\% \quad (1)$$

b. Precision

Precision atau presisi merupakan perhitungan rasio prediksi positif yang benar terhadap semua prediksi positif pada sebuah model yang telah berhasil dibuat, Rumus yang digunakan untuk perhitungan *Precision*, yaitu [10]:

$$\text{Presisi} = \frac{(TP)}{(TP) + (FP)} \times 100\% \quad (2)$$

Keterangan :
True Positif (TP)
False Positif (FP)

c. Recall

Recall adalah rasio prediksi positif yang benar terhadap semua data positif yang sebenarnya pada sebuah model. Rumus yang digunakan untuk perhitungan *Recall*, yaitu [10]:

$$\text{Presisi} = \frac{(TP)}{(TP) + (FN)} \times 100\% \quad (3)$$

Keterangan :
True Positif (TP)
False Negative (FN)

d. F1-Score

F1-Score adalah nilai yang diperoleh dari hasil *Precision* dan *Recall* antara kategori yang diprediksi dengan kategori akurasi.

Tujuan dilakukan perhitungan *F1-Score*, yaitu untuk mengetahui untuk mengetahui keseimbangan nilai pada *precision* dan *recall* [10].

Rumus yang digunakan yaitu :

$$F1 = 2 \cdot \frac{P \times R}{P + R} \quad (4)$$

Keterangan :
Precision (P)
Recall (R)

Di sisi lain, pengujian sistem dilakukan melalui *Black Box Testing* untuk memverifikasi fungsionalitas utama, bertujuan untuk mengetahui data yang dimasukkan akan menghasilkan luaran yang sesuai, serta untuk menemukan kesalahan pada antarmuka, perilaku sistem, dan error handling. Selain itu dilakukan juga mengenai *User Acceptance Testing* (UAT) yang melibatkan operator sebagai pengguna akhir. Penilaian UAT menggunakan kuesioner skala *Likert*



untuk menilai aspek kemudahan penggunaan, kepercayaan terhadap sistem, dan kejelasan klasifikasi yang ditampilkan.

Validitas dan Keandalan

Validitas model dijaga melalui pelabelan manual oleh ahli, serta pengukuran konsistensi menggunakan metode *inter-rater agreement cohent's kappa*, rumus yang digunakan yaitu [16]:

$$\kappa = \frac{P - Pe}{1 - Pe} \quad (5)$$

Keterangan :

Po: Proporsi kesepakatan yang diamati
(observed agreement)

Pe: Proporsi kesepakatan yang diharapkan
secara acak

Selain itu, untuk evaluasi lanjutan dilakukan dengan pendekatan *cross-validation 5-fold* untuk menguji kestabilan prediksi. *cross-validation 5-fold* untuk mengevaluasi performa model dengan cara membagi dataset menjadi beberapa subset (fold), melatih model di sebagian fold, dan menguji di keandalan pada sistem dengan mekanisme pencatatan aktivitas (*logging*), uji beban terbatas untuk mengevaluasi kapasitas sistem, serta desain modular yang memungkinkan pengembangan lebih lanjut secara fleksibel. Rumus yang digunakan untuk perhitungan *cross-validation 5-fold*, yaitu [13]:

$$Cv = \frac{1}{k} \sum_{i=1}^k M_i \quad (6)$$

Keterangan :

k = jumlah fold

Mi = metrik performa (misalnya: akurasi, F1-score, dll) pada fold ke-iii

HASIL DAN PEMBAHASAN

Distribusi Data dan Lingkup Pengujian

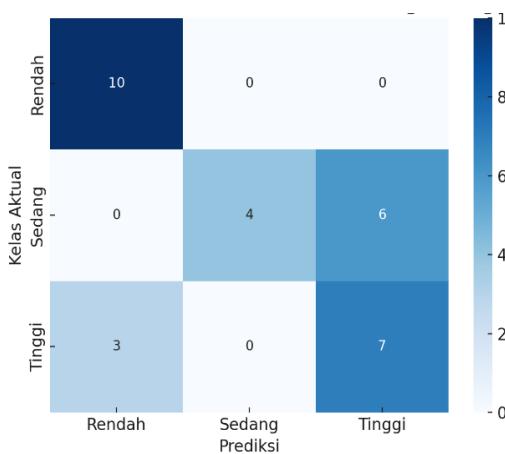
Dataset yang digunakan terdiri dari 150 percakapan teks layanan *hotline* kesehatan mental, masing-masing dilabeli secara manual ke dalam tiga kategori yaitu: rendah, sedang, dan tinggi (masing-masing 50 percakapan pada masing-masing data). Data dibagi menggunakan teknik *stratified split* dengan rasio 80:20 untuk pelatihan dan pengujian model.

Evaluasi Kinerja Model

Model BERT yang di-*fine-tune* pada data pelatihan kemudian dievaluasi pada 30 data uji menggunakan metrik akurasi, *precision*, *recall*, dan F1-score. tabel 1 menunjukkan ringkasan *classification report*, sedangkan gambar 3 menampilkan *confusion matrix*.

Tabel 1. Hasil Evaluasi Model BERT pada Data Uji

Kelas	Precision	Recall	F1-Score	Support
0 (Rendah)	1	1	1	10
1 (Sedang)	0.57	0.4	0.47	10
2 (Tinggi)	0.54	0.7	0.61	10
Accuracy	-	-	-	0.7
Macro avg	0.7	0.7	0.69	30
Weighted avg	0.7	0.7	0.69	30

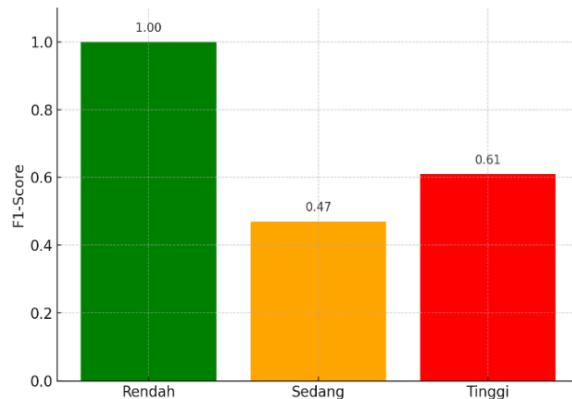


Gambar 3. Confusion matrix hasil klasifikasi model BERT terhadap data uji

Model menunjukkan performa sangat tinggi pada kategori urgensi rendah ($F1 = 1.00$), cukup baik pada tinggi ($F1 = 0.61$), namun rendah pada kategori sedang ($F1 = 0.47$). Hal ini mengindikasikan bahwa model mampu menyaring percakapan non-kritis dengan sangat akurat, namun masih kesulitan membedakan percakapan transisional (ambiguitas antara urgensi rendah dan tinggi).

Visualisasi Performa Kelas

Untuk mempermudah analisis perbandingan setiap antar kelas dapat dilihat pada gambar 3. Visualisasi F1-score model pada masing-masing kelas urgensi dibawah ini, untuk mengetahui grafik batang F1-score dari masing-masing kategori urgensi, yaitu : kategori urgensi rendah ($F1 = 1.00$), cukup baik pada tinggi ($F1 = 0.61$), namun rendah pada kategori sedang ($F1 = 0.47$).



Gambar 4. Visualisasi F1-score model pada masing-masing kelas urgensi

Evaluasi Sistem: Fungsionalitas dan Pengalaman Pengguna

Pengujian sistem dilakukan melalui dua pendekatan:

1. *Black Box Testing* menunjukkan bahwa sistem mampu mengklasifikasikan percakapan dengan akurasi 70% dan menghasilkan prediksi secara *real-time*.
2. UAT dilakukan oleh lima operator *hotline* yang menguji sistem dalam simulasi nyata dapat dilihat pada tabel 2 Hasil evaluasi UAT terhadap pengguna.

Tabel 2. Hasil evaluasi UAT terhadap pengguna

Aspek Evaluasi	Rata-rata
Kemudahan Penggunaan	4.6
Kepercayaan Sistem	4.3
Kejelasan Antarmuka	4.7

Pengguna menyatakan bahwa sistem membantu dalam mempercepat proses klasifikasi dan identifikasi awal percakapan krisis. Namun, mereka juga mencatat keterbatasan sistem dalam menangani ekspresi emosional tersirat atau ambigu.

Pembahasan

Secara umum, sistem klasifikasi urgensi berbasis BERT yang dikembangkan telah menunjukkan kinerja fungsional yang stabil dan tingkat akurasi yang menjanjikan, hal ini dapat dilihat dari hasil akurasi yaitu sebesar 70%



dan F1-score terbaik pada kategori rendah (1.00), cukup baik pada tinggi (0.61), dan lemah pada sedang (0.47). Hasil ini menegaskan potensi NLP dalam mendukung deteksi dini pada layanan psikososial. Keberhasilan dalam mendeteksi kategori urgensi rendah secara sempurna memperlihatkan bahwa model efektif dalam menyaring percakapan non-kritis, sehingga mengurangi beban kerja operator.

Namun, kelemahan pada klasifikasi urgensi sedang dan tinggi menunjukkan perlunya penguatan model terhadap ekspresi krisis yang implisit dan kontekstual.

Kontribusi dan Implikasi

Penelitian ini berkontribusi dalam pengembangan awal sistem deteksi urgensi percakapan dalam konteks layanan kesehatan mental berbasis bahasa Indonesia. Sistem ini telah diuji secara langsung oleh pengguna dan terbukti layak secara teknis maupun praktis. Implikasi implementatif dari sistem ini meliputi:

1. Otomatisasi prioritisasi layanan dalam volume percakapan tinggi
2. Dukungan pengambilan keputusan cepat dalam situasi darurat
3. Potensi integrasi dalam sistem layanan psikososial berbasis teks

Ke depan, penguatan data (terutama pada kelas krisis dan transisi), pengembangan fitur linguistik berbasis emosi, serta eksplorasi pendekatan multimodal (suara, tempo, dan interaktivitas) menjadi arah strategis untuk meningkatkan sensitivitas dan keandalan sistem terhadap kondisi krisis nyata.

KESIMPULAN DAN SARAN

Penelitian ini menghasilkan sistem klasifikasi tingkat urgensi percakapan berbasis BERT yang mampu membedakan tiga kategori urgensi dalam layanan hotline kesehatan mental. Sistem menunjukkan akurasi 70%, dengan F1-score terbaik pada kategori rendah (1.00), cukup baik pada tinggi (0.61), dan lemah pada sedang (0.47). Sistem telah diimplementasikan secara real-time dan mendapatkan umpan balik positif dari operator dalam mempercepat identifikasi dan prioritisasi kasus krisis.

Penelitian selanjutnya disarankan untuk peningkatan jumlah dan keberagaman data, penerapan class weighting dan fine-tuning model, serta integrasi fitur berbasis emosi untuk menangkap sinyal krisis tersirat. Uji validasi lapangan juga penting untuk menilai efektivitas sistem dalam situasi nyata. Aspek etika dan privasi data perlu dijaga ketat. Dengan penguatan tersebut, sistem ini berpotensi menjadi alat bantu krusial dalam layanan psikososial digital di Indonesia..

DAFTAR PUSTAKA

- [1] Alasmari, A. 2025. "A scoping review of Arabic Natural Language Processing for mental health." Healthcare, vol. 13, no. 9, hh. 963. doi: 10.3390/healthcare13090963.
- [2] Anandan, R., Cross, W., Huy, N., dan Olasoji, M. 2024. "Mental health nurses' empathy towards consumers with dual diagnosis: a descriptive study." J. Clin. Nurs., vol. 33, no. 8, hh. 3199–3211. doi: 10.1111/jocn.17011.
- [3] Areshay, A., dan Mathkour, H. 2023. "Transfer learning for sentiment classification using bidirectional encoder representations from transformers (BERT) model." Sensors, vol. 23, no. 11, hh. 5232. doi: 10.3390/s23115232.
- [4] Baldaçara, L., et al. 2023. "Brazilian psychiatric association guidelines for the management of suicidal behavior. Part 3. Suicide prevention hotlines." Braz. J. Psychiatry. doi: 10.4762/1516-4446-2022-2536.
- [5] Chandran, D., et al. 2019. "Use of natural language processing to identify obsessive compulsive symptoms in patients with schizophrenia, schizoaffective disorder or bipolar disorder." Sci. Rep., vol. 9. doi: 10.1038/s41598-019-49165-2.
- [6] Glaz, R., Haralambous, Y., Kim-Dufor, D., Lenca, P., Billot, R., Ryan, T., et al. 2021. "Machine learning and natural language processing in mental health: systematic review." J. Med. Internet Res., vol. 23, no. 5, hh. e15708. doi: 10.2196/15708.
- [7] Heinz, I., Mergl, R., Hegerl, U., Rummel-Kluge, C., dan Kohls, E. 2019. "Depression stigma and management of suicidal callers: a cross-sectional survey of crisis hotline counselors." BMC Psychiatry, vol. 19, no. 1. doi: 10.1186/s12888-019-2325-y.
- [8] Isni, K., et al. 2024. "The implementation of adolescent mental health services: opportunities and challenges." Int. J. Public Health Sci., vol. 13, no. 2, hh. 810. doi: 10.11591/ijphs.v13i2.23597.
- [9] Kotera, Y., dan Gordon, W. 2019. "Japanese managers' experiences of neuro-linguistic programming: a qualitative investigation." J. Ment. Health Train. Educ. Pract., vol. 14, no. 3, hh. 174–185. doi: 10.1108/jmtep-06-2018-0033.
- [10] Liu, Z., et al. 2021. "Listening to mental health crisis needs at scale: using natural language processing to understand and evaluate a mental health crisis text messaging service." Front. Digit. Health, vol. 3. doi: 10.3389/fdgh.2021.779091.
- [11] Namburi, S., dan Hopkins, G. 2024. "Beyond content: a trauma-informed framework for academic writing evaluation." ICER, vol. 1, no. 1, hh. 227–237. doi: 10.34190/icer.1.1.3112.
- [12] Richter, A., Sjunnestrand, M., Strandh, M., dan Hasson, H. 2022. "Implementing school-based mental health services: a scoping review of the literature summarizing the factors that affect implementation." Int. J. Environ. Res. Public Health, vol. 19, no. 6, hh. 3489. doi: 10.3390/ijerph19063489.
- [13] Ridgway, J., et al. 2021. "Natural language processing of clinical notes to identify mental illness and substance use among people living with HIV: retrospective cohort study." JMIR Med. Inform., vol. 9, no. 3, hh. e23456. doi: 10.2196/23456.
- [14] Saha, S., et al. 2021. "Unique collaboration of modern medicine and traditional faith-healing for the treatment of mental illness." J. Family Med. Prim. Care, vol. 10, no. 1, hh. 521–526. doi: 10.4103/jfmpc.jfmpc_979_19.



- [15] Scherbakov, D., Hubig, N., Lenert, L., Alekseyenko, A., dan Obeid, J. 2025. "Natural language processing and social determinants of health in mental health research: AI-assisted scoping review." *JMIR Ment. Health*, vol. 12, hh. e67192. doi: 10.2196/67192.
- [16] Schopow, N., Osterhoff, G., dan Baur, D. 2023. "Applications of the natural language processing tool ChatGPT in clinical practice: comparative study and augmented systematic review." *JMIR Med. Inform.*, vol. 11, hh. e48933. doi: 10.2196/48933.
- [17] Suhartono, D., Saputra, I., Pratama, A., dan Nathaniel, G. 2024. "Psychological stress detection using transformer-based models." *ComTech*, vol. 15, no. 1, hh. 65–71. doi: 10.21512/comtech.v15i1.11105.
- [18] Teferra, B., et al. 2024. "Screening for depression using natural language processing: literature review." *Interact. J. Med. Res.*, vol. 13, hh. e55067. doi: 10.2196/55067.
- [19] Tong, Y., et al. 2020. "Prospective study of association of characteristics of hotline psychological intervention in 778 high-risk callers with subsequent suicidal act." *Aust. N. Z. J. Psychiatry*, vol. 54, no. 12, hh. 1182–1191. doi: 10.1177/0004867420963739.
- [20] Viani, N., et al. 2021. "A natural language processing approach for identifying temporal disease onset information from mental healthcare text." *Sci. Rep.*, vol. 11. doi: 10.1038/s41598-020-80457-0.