Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi
https://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/index
Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru
Loktabat – Banjarbaru (Tlp. 0511 4782881), e-mail: puslit.stmikbjb@gmail.com
e-ISSN: 2685-0893

Pengembangan Model *Deep Learning* dengan *Slang-Aware Embeddings* untuk Deteksi Promosi Judi *Online*

DOI: http://dx.doi.org/10.35889/jutisi.v14i3.3226

Yo'el Pieter Sumihar[1*], Kristian Juri Damai Lase[2], Juli Herman Lase[3]
Informatika, Universitas Kristen Immanuel, Yogykarta, Indonesia
*e-mail *Corresponding Author*: pieter.haro@ukrimuniversity.ac.id

Abstract

Online gambling promotions on social media platforms such as YouTube often employ slang or non-standard language to evade traditional moderation systems, increasing the spread of illegal content in Indonesia. This study aims to develop a hybrid deep learning model combining BERT and LSTM to accurately detect online gambling promotions. Data were collected from YouTube comments through a web scraping process, followed by text cleaning, labeling, and normalization using a semi-automatic slang dictionary. The model was trained with slang-aware embeddings to capture informal language context. Evaluation was conducted using precision, recall, F1-score, and confusion matrix metrics. The results show an accuracy of 96% with an F1-score of 0.96, indicating a strong balance between precision and recall. These findings demonstrate the effectiveness of the proposed hybrid approach in automatically detecting online gambling promotional content.

Kata kunci: *Online Gambling Detection; Deep Learning; NLP; Slang-Aware Embeddings; BERT-LTSM*

Abstrak

Promosi judi daring di media sosial seperti YouTube sering menggunakan bahasa tidak baku atau slang untuk menghindari deteksi sistem moderasi. Kondisi ini berpotensi meningkatkan penyebaran konten ilegal di Indonesia. Penelitian ini bertujuan mengembangkan model deep learning hibrida yang mengombinasikan BERT dan LSTM guna mendeteksi promosi judi daring secara lebih akurat. Data dikumpulkan dari komentar YouTube melalui proses *web scraping*, kemudian diproses melalui tahap pembersihan teks, pelabelan, dan normalisasi menggunakan kamus slang semi-otomatis. Model dilatih dengan *slang-aware embeddings* untuk menangkap konteks bahasa tidak resmi. Pengujian dilakukan menggunakan metrik *precision, recall*, F1-*score*, dan *confusion matrix*. Hasil menunjukkan akurasi sebesar 96% dengan nilai F1-*score* 0,96, menandakan keseimbangan tinggi antara presisi dan sensitivitas model. Temuan ini membuktikan efektivitas pendekatan hibrida dalam mendeteksi konten promosi judi daring secara otomatis.

Kata kunci: *Deteksi Judi Online; Deep Learning; NLP; Slang-Aware Embeddings; BERT - LTSM*

1. Pendahuluan

Perjudian daring telah menjadi fenomena yang marak, terutama di media sosial. Para *promotor* judi sering kali menggunakan bahasa tidak resmi atau *slang*, seperti "WD" (*withdraw*) dan "depo" (*deposit*), untuk menghindari sistem moderasi konvensional. Di Indonesia, di mana praktik ini ilegal, penyebaran konten judi daring menimbulkan dampak sosial dan ekonomi yang signifikan, terutama bagi remaja. Oleh karena itu, dibutuhkan sistem deteksi cerdas yang mampu mengenali pola bahasa informal tersebut untuk mengatasi masalah ini secara efektif.

Pendekatan *deep learning* telah terbukti efektif dalam berbagai tugas pemrosesan bahasa alami. Studi-studi sebelumnya menunjukkan bahwa model *LSTM* mampu mengklasifikasikan ujaran kebencian dengan akurasi tinggi hingga 94,66% [1]. Sementara itu, penelitian lain menunjukkan bahwa model *BERT* memiliki performa lebih unggul dari *LSTM* dalam

analisis sentimen terkait Pemilu 2024 [2]. Selain itu, studi yang menggunakan *RoBERTa* dalam analisis sentimen di YouTube juga menghasilkan nilai positif 93% [3].

Dalam konteks bahasa Indonesia, penelitian tentang analisis sentimen ulasan TikTok menggunakan *IndoBERTweet* dan *LSTM* menunjukkan efektivitasnya dalam memahami bahasa gaul [4]. Studi lain yang membandingkan *LSTM* dan *IndoBERT* untuk mendeteksi hoaks di Twitter menemukan bahwa kombinasi keduanya dapat meningkatkan akurasi deteksi [5]. Keberhasilan model-model ini menunjukkan potensi besar untuk diterapkan dalam deteksi promosi judi daring. Beberapa penelitian juga telah mengembangkan metode deteksi konten ilegal, seperti optimalisasi bot Telegram untuk mendeteksi situs judi [6] dan implementasi *BiLSTM* untuk mendeteksi potensi depresi dari unggahan media sosial [7], yang membuktikan bahwa pendekatan deep learning efektif dalam mengatasi masalah ini.

Berdasarkan tantangan yang ada serta keberhasilan metode sebelumnya, penelitian ini bertujuan mengembangkan model *deep learning* hibrida yang menggabungkan *BERT* dan *LSTM*. Model ini akan diperkuat dengan *embeddings* yang peka terhadap *slang* (*slang-aware embeddings*) untuk menangkap istilah tidak resmi yang sering digunakan dalam promosi judi *online*. Penelitian ini juga akan menghasilkan kamus istilah *slang* judi daring yang dapat diakses publik serta API untuk implementasi *real-time*. Dengan demikian, penelitian ini dapat memberikan kontribusi nyata dalam mendukung upaya pemerintah dan platform digital dalam mengatasi peredaran konten perjudian ilegal.

2. Tinjauan Pustaka

Tinjauan pustaka kami menunjukkan bahwa penelitian mengenai deteksi konten promosi judi daring di media sosial masih sangat minim. Oleh karena itu, untuk mengatasi keterbatasan ini, kami mengumpulkan pengetahuan dari studi-studi terkait. Kami berfokus pada literatur yang membahas karakteristik konten serupa serta teknik yang digunakan dalam deteksi *spam*, klasifikasi teks, dan identifikasi konten berbahaya.

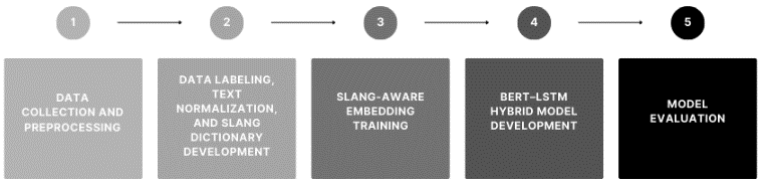
Eksplorasi kami terhadap perjudian daring memberikan wawasan mengenai pola tekstual dalam iklan judi. Meskipun studi [8] dan [9] kurang menyertakan evaluasi mendalam, studi [8] menemukan bahwa iklan judi sering menggunakan bahasa positif yang menyoroti bonus dan penawaran khusus, sebuah tren yang juga diamati pada studi [10],[11], [12], [13] .Studi [10] secara khusus mengidentifikasi karakteristik iklan judi di Indonesia, seperti menonjolkan keuntungan finansial dan kemudahan bergabung. Temuan ini menegaskan kelangkaan penelitian tentang deteksi konten judi daring, khususnya di Indonesia, sehingga menekankan pentingnya pendekatan komprehensif kami untuk memahami strategi promosi tersebut.

Dalam konteks klasifikasi *spam*, khususnya promosi judi, banyak studi telah membandingkan berbagai algoritma dan prapemrosesan. Studi [14] dan [15] merekomendasikan algoritma SVM (*Support Vector Machine*) untuk deteksi *spam*, dengan studi [15] berhasil mencapai *F-Score* 96% menggunakan prapemrosesan *bi-gram* dan *TF-IDF*. Pendekatan lain juga menunjukkan hasil yang kuat: studi [16] menggunakan turunan *Thai BERT* dengan *F-Score* 0,8, sementara studi [17] dan [18] melaporkan *F-Score* sangat tinggi, masing-masing 99,73% dan 93%, melalui kombinasi jaringan saraf dengan *GloVe*, dan regresi logistik dengan *Word2Vec*. Selain itu, studi [19] menyajikan pendekatan berbasis aturan *regex*, meskipun evaluasinya kurang terperinci.

Berbagai penelitian sebelumnya telah menunjukkan efektivitas model *BERT*, *LSTM*, dan turunannya dalam tugas analisis sentimen serta deteksi ujaran kebencian pada teks berbahasa Indonesia. Meskipun demikian, sebagian besar studi tersebut masih berfokus pada teks formal atau bahasa standar, sementara karakteristik bahasa yang digunakan dalam promosi judi daring di Indonesia memiliki ciri khas tersendiri berupa penggunaan *slang*, simbol, serta modifikasi huruf yang bertujuan untuk menghindari sistem deteksi otomatis. Untuk menjawab celah tersebut, penelitian ini menawarkan pendekatan baru melalui integrasi model *BERT–LSTM dengan slang-aware embeddings*, yang didukung oleh kamus slang semi-otomatis yang dikembangkan secara kontekstual dari data komentar asli di YouTube. Pendekatan ini menjadikan model lebih adaptif terhadap dinamika bahasa informal yang berkembang di media sosial, sekaligus memberikan kontribusi nyata terhadap pengembangan sistem deteksi otomatis konten ilegal berbasis bahasa Indonesia yang lebih kontekstual dan responsif terhadap variasi linguistik pengguna.

3. Metodologi

Tahapan penelitian ini disusun secara sistematis untuk memastikan setiap prosesnya mencapai tujuan yang ditetapkan. Langkah-langkah tersebut mencakup pengumpulan dan pemrosesan data awal, pengembangan sumber daya linguistik berupa kamus *slang*, pelatihan *embeddings* yang peka terhadap *slang*, perancangan model klasifikasi berbasis *deep learning*, dan evaluasi komprehensif terhadap kinerja model. Tahapan ini ditunjukkan pada Gambar 1 dan dijelaskan secara rinci sebagai berikut:



Gambar 1. Langkah Metodologi Penelitian

3.1. Pengumpulan dan Pra-pemrosesan Data

Data dikumpulkan dari komentar-komentar YouTube yang mengindikasikan promosi judi daring melalui proses *web scraping* otomatis. Data yang diperoleh kemudian dibersihkan dari berbagai elemen yang tidak relevan, seperti simbol, tautan, dan elemen non-teks, menggunakan prosedur pra-pemrosesan standar.

3.2. Pelabelan Data, Normalisasi Teks, dan Pengembangan Kamus Slang

Setiap item data dilabeli secara semi-otomatis dengan mengidentifikasi kata atau frasa yang mengindikasikan promosi judi daring. Verifikasi manual juga dilakukan untuk meninjau istilah yang tidak dikenali oleh sistem, sekaligus mengembangkan kamus *slang*. Kamus ini berfungsi untuk mengonversi istilah slang ke dalam bentuk standar yang dapat dipahami oleh model.

3.3. Pelatihan Embedding yang Peka terhadap Slang

Untuk meningkatkan pemahaman model terhadap konteks istilah *slang*, *custom embeddings* dibuat menggunakan *Word2Vec*. Kalimat-kalimat yang telah di-tokenisasi dari data teks yang bersih digunakan untuk melatih model *Word2Vec*, yang menghasilkan *word embeddings* yang mampu menangkap hubungan semantik antar istilah *slang*. Sebuah fungsi diimplementasikan untuk merata-ratakan embeddings Word2Vec dari semua kata dalam kalimat, sehingga setiap kalimat dapat direpresentasikan sebagai vektor berukuran tetap. Vektor kalimat yang dihasilkan kemudian digunakan sebagai fitur untuk model, memungkinkannya untuk memahami konteks kata-kata slang dalam promosi judi daring dengan lebih baik.

3.4. Pengembangan Model Hibrida BERT-LSTM

Model hibrida ini mengombinasikan kemampuan *BERT* dalam mengekstrak representasi semantik kata dengan kekuatan *LSTM* dalam menangkap urutan dan konteks historis. Vektor fitur yang dihasilkan oleh *BERT* kemudian diteruskan ke *LSTM* untuk proses klasifikasi.

3.5. Evaluasi Model

Kinerja model dievaluasi menggunakan metrik presisi, *recall*, dan *F1-score*. Model ini kemudian diuji pada data baru yang tidak termasuk dalam data pelatihan atau evaluasi untuk menilai kemampuan generalisasinya.

4. Hasil dan Pembahasan

4.1 Pengumpulan dan Pra-pemrosesan Data

Proses pengumpulan data dilakukan melalui *web scraping* komentar YouTube yang berpotensi mengandung promosi judi daring. Kata kunci yang digunakan antara lain *“slot”, “toto”, “gacor”, “jepe”,* dan *“deposit”*. Dari hasil *scraping* diperoleh sekitar 10.000 komentar mentah dari berbagai video.

Data yang dikumpulkan dari komentar YouTube, beserta hasil dari langkah-langkah pra-pemrosesan, disajikan dalam Tabel 1. Tabel ini memberikan gambaran komprehensif mengenai

komentar mentah yang dikumpulkan selama fase pengumpulan data, serta transformasi yang diterapkan pada komentar-komentar tersebut selama pra-pemrosesan.

Tabel 1. Contoh Dataset

Text Asli	Text Bersih
Di jamin JP dengan sedekah seikhlasnya	di jamin jp dengan sedekah seikhlasnya
gachor banget di AERO88, jepey gmpng	gachor banget di aero88 jepey gmpng
#DEWA DORA# gampang banget buat WD, rugi kalo gak coba	dewa dora gampang banget buat wd rugi kalo gak coba

4.2 Pelabelan Data dan Pengembangan Kamus Slang

Setiap komentar diberi label promosi (1) atau non-promosi (0) berdasarkan kemunculan istilah yang mengindikasikan ajakan bermain judi atau penyebutan situs tertentu. Proses ini dilakukan secara semi-otomatis menggunakan daftar kata kunci dan diverifikasi manual oleh peneliti dan asisten mahasiswa.

Bersamaan dengan itu, dikembangkan kamus *slang* semi-otomatis yang memetakan kata tidak baku ke bentuk standarnya, misalnya:
"jepe" → "jackpot", "g4c0rr" → "gacor", "wedeh" → "withdraw".
Kamus ini digunakan dalam tahap normalisasi untuk meningkatkan konsistensi data dan membantu model memahami variasi penulisan.

Contoh hasil pelabelan data dan normalisasi teks disajikan pada Tabel 2, yang mengilustrasikan bagaimana setiap komentar dikategorikan dan distandarisasi. Selain itu, Tabel 3 memuat contoh dari kamus slang, menampilkan istilah-istilah *slang* yang teridentifikasi beserta bentuk standarnya yang digunakan untuk meningkatkan pemahaman model terhadap teks.

Tabel 2. Contoh Data yang Telah Dilabeli dan Dinormalisasi

Text Bersih	Label	Text Normalisasi
di jamin jp dengan sedekah seikhlasnya	1	di jamin Jackpot dengan sedekah seikhlasnya
gachor banget di aero88 jepey gmpng	1	Gacor banget di aero88 Jackpot gmpng
dewa dora gampang banget buat wd rugi kalo gak coba	1	dewa dora gampang banget buat Withdraw rugi kalo gak coba
episode ini gilaaak bangetttttt full ngakak	0	episode ini gilaaak bangetttttt full ngakak

Contoh hasil pelabelan dan normalisasi teks disajikan pada Tabel 2. Tabel ini menunjukkan bagaimana komentar mentah dikategorikan sebagai promosi judi daring atau bukan, dan bagaimana normalisasi diterapkan untuk menstandarkan konten. Misalnya, istilah slang seperti "di jamin jp" dinormalisasi menjadi "di jamin *Jackpot*," dan *"gachor"* dikoreksi menjadi *"Gacor"*. Hal ini bertujuan membuat teks lebih konsisten dan mudah dipahami oleh model. Selain itu, komentar yang tidak terkait dengan promosi judi, seperti "episode ini gilaaak bangetttttt full ngakak," diberi label sebagai contoh negatif (0).

Tabel 3. Mengembalikan kata slang ke bentuk normal

Slang	Kata Standart
Jepe	*Jackpot*
g4c0rr	Gacor
g4c0rrr	Gacor
Wedeh	*Withdraw*
jeipe	*Jackpot*

Tabel 3 menyajikan contoh-contoh dari kamus *slang*. Tabel ini menampilkan beragam istilah *slang* yang ditemukan dalam himpunan data dan bentuk standarnya. Istilah *slang* ini, yang sering digunakan dalam promosi judi daring, dipetakan ke padanan yang lebih formal atau mudah dikenali. Sebagai contoh, istilah "Jepe" distandarkan menjadi *"Jackpot,"* dan *"g4c0rr"* dikoreksi menjadi *"Gacor."* Entri lain, seperti *"Wedeh"* dan *"jeipe,"* dipetakan ke kata *"Withdraw,"* sehingga memastikan model dapat menafsirkan istilah-istilah ini secara konsisten selama analisis.

4.3 Pelatihan *Slang-Aware Embeddings*

Tahap pelatihan *slang-aware embeddings* dilakukan untuk memetakan hubungan semantik antar kata dalam korpus komentar YouTube yang mengandung indikasi promosi judi daring. Dataset yang digunakan telah melalui proses pembersihan, normalisasi, dan konversi *slang* berdasarkan *slang dictionary* yang dikembangkan dari data aktual. Setiap kata *slang* seperti *"gacor", "jepei", "slot",* dan *"maxwin"* dikonversi ke bentuk representatif yang tetap mempertahankan makna kontekstualnya, sehingga model dapat mengenali pola bahasa tidak baku yang sering digunakan oleh pelaku promosi.

Pelatihan dilakukan menggunakan algoritma *Word2Vec* dengan parameter: *vector size* = 100, *window size* = 5, *min_count* = 1, dan *training algorithm* = *skip-gram* (sg=1). Model dilatih pada token hasil pemrosesan teks yang telah dinormalisasi menggunakan kamus *slang*, menghasilkan vektor representasi berdimensi 100 untuk setiap kata unik.

Untuk mengevaluasi hasil pelatihan, dilakukan uji *most similar words* untuk sejumlah istilah *slang* yang sering muncul dalam konten promosi judi daring. Lima kata terdekat dengan setiap istilah slang dianalisis berdasarkan nilai *cosine similarity*, yang menunjukkan kedekatan semantik antar kata dalam ruang vektor.

Tabel 4. Hasil Pelatihan Slang-Aware Embeddings

Kata Slang	Top-5 Kata Terdekat	Cosine Similarity Rata-rata
gacor	mcb168, abis, gmpng, gila, mulu	0.943
jepei	sorjp88, rejeki, cair, pokoknya, mendadak	0.961
hoki	terusterusan, ngocor, nyampe, melulu, turun	0.960
slot	g4c0rnya, ngasih, dibet4d, gawe212, joss	0.948
menang	mulu, gac0r, jeipenya, gac0rnya, s3l0tnya	0.940
toto	bebas, dibolehkan, dibilang, 18, lambung	0.997
withdraw	member, aja, jamin, lngsng, garansi	0.832

Hasil pada Tabel 4 menunjukkan bahwa model *Word2Vec* berhasil menangkap hubungan semantik yang konsisten antara istilah slang dengan kata yang muncul pada konteks promosi. Misalnya, kata *"gacor"* memiliki kedekatan tinggi dengan "mcb168" dan *"slot",* yang merupakan nama situs dan istilah kemenangan. Demikian pula, *"jepei"* berhubungan erat dengan "rejeki" dan "cair", merepresentasikan konteks kemenangan atau pencairan hadiah. Nilai *cosine similarity* rata-rata yang tinggi (0,94–0,99) menunjukkan stabilitas representasi semantik yang dihasilkan oleh model.

Sementara itu, beberapa istilah seperti *"toto"* dan *"withdraw"* menunjukkan konteks campuran, karena digunakan baik dalam konteks promosi maupun non-promosi. Hal ini menunjukkan bahwa meskipun model mampu mengenali asosiasi semantik utama, beberapa istilah masih memiliki makna ganda yang memerlukan interpretasi kontekstual lebih lanjut.

Secara keseluruhan, hasil pelatihan ini membuktikan bahwa *slang-aware embeddings* efektif dalam mempelajari hubungan semantik dari teks informal dan kreatif yang sering digunakan dalam aktivitas promosi judi daring. Model ini kemudian digunakan sebagai lapisan representasi awal pada arsitektur *hybrid deep learning BERT–LSTM*.

4.4 Pengembangan Model Hibrida BERT–LSTM

Pada fase ini dilakukan integrasi antara *BERT (Bidirectional Encoder Representations from Transformers)* yang berfungsi mengekstraksi representasi semantik teks dengan *LSTM (Long Short-Term Memory)* yang mampu menangkap hubungan kontekstual secara sekuensial. Arsitektur model dirancang dengan mengalirkan keluaran (*output*) dari lapisan *BERT* ke lapisan *LSTM*, kemudian diteruskan ke lapisan dense untuk melakukan klasifikasi biner, yaitu komentar

promosi (kelas 1) dan non-promosi (kelas 0). Proses pelatihan dilakukan dengan pembagian data sebesar 80% untuk *training* dan 20% untuk *testing*.

Hasil *fine-tuning* menunjukkan performa model yang tinggi dan seimbang, dengan akurasi 96%, *precision* 0,95, *recall* 0,97, dan *F1-score* 0,96. Evaluasi menggunakan metrik *presisi*, *recall*, dan *F1-score* menunjukkan kemampuan model dalam mengidentifikasi komentar promosi judi daring dengan akurat sekaligus membedakannya dari konten non-promosi. Hasil evaluasi kinerja model untuk kedua kelas dapat di lihat pada Table 5.

Tabel 5. Hasil evaluasi Model Deep Learning

Metric	Class 0 (Non-Promosi)	Class 1 (Promosi)
Precision	0.95	0.97
Recall	0.97	0.95
F1-Score	0.96	0.96

Nilai *F1-score* yang identik (0,96) pada kedua kelas menunjukkan bahwa model memiliki keseimbangan yang baik antara tingkat ketepatan dan kemampuan deteksi, serta efektivitas tinggi dalam membedakan komentar promosi dan non-promosi secara konsisten.

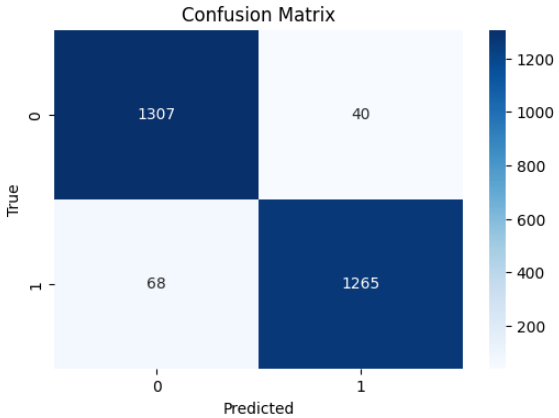
4.5 Evaluasi Model dan Analisis Hasil

Evaluasi model menggunakan *confusion matrix* menunjukkan jumlah kesalahan klasifikasi yang sangat rendah, yaitu 68 *false positives* dan 40 false *negatives* dari total ribuan data uji. Hasil ini menandakan bahwa model mampu mengenali pola bahasa promosi judi daring secara akurat meskipun terdapat variasi penulisan *slang* dan penggunaan karakter non-baku.

Gambar 2 menampilkan *confusion matrix* yang menggambarkan kemampuan model dalam membedakan komentar promosi (*Class 1*) dan non-promosi (*Class 0*). Rincian hasil klasifikasi adalah sebagai berikut:

1. *True Positive* (TP) untuk *Class 0* berjumlah 1.307, menunjukkan komentar non-promosi yang teridentifikasi dengan benar.
2. *False Negative* (FN) untuk *Class 0* berjumlah 40, yaitu komentar non-promosi yang keliru diklasifikasikan sebagai promosi.
3. *False Positive* (FP) untuk *Class 1* berjumlah 68, yaitu komentar promosi yang salah dikenali sebagai non-promosi.
4. *True Positive* (TP) untuk *Class 1* berjumlah 1.265, menunjukkan komentar promosi yang berhasil diklasifikasikan secara tepat.

Secara keseluruhan, model menunjukkan tingkat akurasi yang sangat tinggi dengan kesalahan minimal. Nilai TP yang besar serta FP dan FN yang rendah membuktikan bahwa model hibrida *BERT–LSTM* dengan *slang-aware embeddings* mampu mengklasifikasikan komentar promosi dan non-promosi secara konsisten dan andal, bahkan pada data yang memiliki ragam penulisan *slang*.



Gambar 2. *Confusion Matrix*

4.6 Pembahasan

Hasil penelitian ini menunjukkan bahwa integrasi *slang-aware embeddings* dengan arsitektur *BERT–LSTM* mampu meningkatkan performa deteksi konten promosi judi daring secara signifikan. Model yang dikembangkan mencapai accuracy sebesar 96% dan *F1-score* sebesar 0.96 dengan keseimbangan antara *false positive* dan *false negative*, menunjukkan kemampuan generalisasi yang baik terhadap komentar baru di YouTube. Temuan ini memperkuat hasil penelitian sebelumnya [1] yang membuktikan efektivitas jaringan *LSTM* dalam memahami konteks ujaran berbahasa Indonesia, namun penelitian tersebut masih terbatas pada ujaran kebencian dengan bahasa formal. Pendekatan dalam penelitian ini memperluas penerapan *LSTM* pada bahasa informal dengan menambahkan lapisan *slang normalization* dan *domain-specific embeddings* yang mampu mengenali variasi ejaan kreatif seperti g4c0r dan s1ot yang sering digunakan untuk menghindari sistem deteksi otomatis.

Kinerja model yang tinggi juga sejalan dengan penelitian sebelumnya [2], [3], [4] yang menunjukkan bahwa kombinasi transformer seperti *BERT* dan *LSTM* dapat meningkatkan hasil analisis sentimen pada teks berbahasa Indonesia. Namun, berbeda dengan penelitian-penelitian tersebut yang berfokus pada teks formal, penelitian ini menambahkan kontribusi baru dengan memperkuat kemampuan model dalam memahami bahasa tidak baku melalui pelatihan *slang-aware embeddings*. Dengan demikian, penelitian ini tidak hanya memperkuat arah pengembangan model kontekstual sebagaimana dijelaskan dalam [5], tetapi juga memperkaya pemahaman bahwa integrasi antara *contextual embeddings* dan *domain-adaptive embeddings* menghasilkan representasi semantik yang lebih stabil pada teks dengan variasi ejaan dan gaya bahasa ekstrem.

Selain memperkuat penelitian NLP sebelumnya, hasil penelitian ini juga melengkapi pendekatan praktis yang dikembangkan dalam [6], yang membangun sistem bot Telegram untuk mendeteksi situs perjudian daring. Sementara penelitian [6] berfokus pada aspek implementasi sistem, penelitian ini memberikan penguatan pada aspek kecerdasan linguistik yang dapat diintegrasikan langsung ke dalam sistem tersebut. Dengan demikian, kedua pendekatan ini saling melengkapi antara sisi rekayasa perangkat lunak dan sisi pemodelan bahasa berbasis AI.

Dari perspektif global, hasil penelitian ini konsisten dengan temuan sebelumnya [8] yang menyoroti strategi penyamaran promosi judi daring di media sosial melalui narasi positif dan penggunaan bahasa komunitas. Fenomena serupa juga ditemukan dalam korpus komentar Indonesia, menunjukkan bahwa pola linguistik promosi bersifat lintas-bahasa. Dengan demikian, pendekatan *slang-aware embeddings* dalam penelitian ini berpotensi diadaptasi untuk deteksi konten sejenis di bahasa lain. Selain itu, temuan ini juga berkaitan dengan penelitian [14], [16], [17], yang menerapkan *deep learning* untuk deteksi *spam* dan *phishing*. Namun, penelitian-penelitian tersebut dilakukan pada bahasa Inggris atau Arab, sedangkan penelitian ini memperluas konteks ke bahasa Indonesia dengan karakteristik sosial-linguistik yang berbeda, terutama dalam aspek penggunaan huruf campuran, singkatan, dan variasi bunyi.

Secara sosial dan kebijakan, hasil penelitian ini mendukung pandangan yang disampaikan dalam [10], [11] yang menyoroti pentingnya tanggung jawab hukum terhadap pelaku promosi judi daring, serta membahas konvergensi antara *gaming* dan *gambling* dalam ekosistem digital. Pendekatan deteksi otomatis yang diusulkan dalam penelitian ini dapat menjadi dukungan teknis untuk upaya pencegahan penyebaran konten ilegal di ruang digital sebagaimana dibahas pada penelitian-penelitian tersebut.

Secara ilmiah, penelitian ini menyatukan dua arah besar dalam pengembangan *Natural Language Processing* (NLP), yaitu *contextual embeddings* yang berfokus pada pemahaman konteks linguistik dan *domain-specific embeddings* yang menyesuaikan makna dengan gaya bahasa pengguna. Integrasi keduanya menghasilkan model yang adaptif terhadap teks tidak baku dan dapat menangkap konteks promosi yang tersembunyi di balik pola ejaan kreatif. Temuan ini memperkaya khazanah riset NLP berbahasa Indonesia yang sebelumnya berfokus pada bahasa formal, serta memberikan landasan konseptual bagi pengembangan sistem moderasi konten berbasis AI yang mampu beradaptasi terhadap dinamika bahasa di media sosial. Dengan demikian, penelitian ini tidak hanya mempertegas validitas hasil-hasil sebelumnya tetapi juga memberikan penguatan metodologis yang signifikan bagi pengembangan model deteksi konten berisiko sosial di Indonesia.

5. Simpulan

Berdasarkan hasil penelitian dan evaluasi model yang dilakukan, dapat disimpulkan bahwa pendekatan *slang-aware embeddings* yang diintegrasikan dengan arsitektur *BERT–LSTM* terbukti efektif dalam mendeteksi konten promosi judi daring pada komentar YouTube berbahasa Indonesia. Melalui proses pelatihan pada data komentar yang telah dinormalisasi menggunakan kamus *slang*, model berhasil memahami hubungan semantik antara istilah-istilah tidak baku seperti "gacor", "slot", dan "jepei" dengan konteks kemenangan, situs promosi, serta istilah umum yang digunakan oleh pelaku promosi judi daring. Hasil pelatihan embedding menunjukkan nilai *cosine similarity* yang tinggi, yaitu berkisar antara 0,83 hingga 0,99, yang menandakan bahwa model *Word2Vec* yang digunakan mampu menangkap makna kontekstual *slang* dengan stabil dan representatif.

Berdasarkan hasil evaluasi melalui matriks kebingungan, model yang dikembangkan menunjukkan performa yang sangat baik. Model berhasil mengklasifikasikan sebagian besar komentar dengan tepat, di mana jumlah *True Positive* mencapai 1.307 untuk kelas non-promosi dan 1.265 untuk kelas promosi. Kesalahan klasifikasi juga sangat rendah, hanya ditemukan 68 komentar promosi yang salah diklasifikasikan sebagai non-promosi (*False Positive*) dan 40 komentar non-promosi yang salah diklasifikasikan sebagai promosi (*False Negative*). Dengan nilai *F1-score* sebesar 0,96 pada kedua kelas dan akurasi keseluruhan mencapai 96%, model menunjukkan keseimbangan yang baik antara presisi dan recall. Kinerja ini membuktikan bahwa integrasi *slang-aware embeddings* memberikan kontribusi nyata terhadap peningkatan sensitivitas model terhadap bahasa informal dan ejaan kreatif yang sering digunakan dalam aktivitas promosi ilegal di media sosial. Secara keseluruhan, pendekatan yang dikembangkan dalam penelitian ini menghasilkan model deteksi yang akurat, stabil, dan adaptif, sehingga dapat berfungsi sebagai alat bantu yang efektif dalam upaya moderasi dan pencegahan penyebaran promosi judi daring di platform digital seperti YouTube.

Ucapan Terimakasih

Penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada Hibah Penelitian dari Kemdiktisaintek dan LPPM Universitas Kristen Immanuel atas pendanaan penelitian ini. Tanpa dukungan tersebut, riset ini tidak mungkin terlaksana.

Daftar Referensi

[1] A. Perwira and J. Dwitama, "Deteksi Ujaran Kebencian pada Teks Bahasa Indonesia Menggunakan Bidirectional Long Short Term Memory (Bi-LSTM)," Universitas Islam Indonesia, Yogyakarta, 2023.

[2] M. Khadapi and V. Maruli Pakpahan, "Analisis Sentimen Berbasis Jaringan LSTM dan BERT terhadap Diskusi Twitter tentang Pemilu 2024," *JUKI : Jurnal Komputer dan Informatika*, vol. 6, no. 2, pp. 130–137, Nov. 2024.

[3] Y. P. Sumihar, "Sentiment Analysis of Public Opinions Regarding 'Ideas of Presidential Candidates' in YouTube Video Comments with Robustly Optimized BERT Pretraining Approach," *Jurnal Sistem Informasi dan Ilmu Komputer Prima*, vol. 8, no. 1, pp. 12–28, Aug. 2024.

[4] J. C. Setiawan, K. M. Lhaksmana, and B. Bunyamin, "Sentiment Analysis of Indonesian TikTok Review Using LSTM and IndoBERTweet Algorithm," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 3, pp. 774–780, Aug. 2023, doi: 10.29100/jipi.v8i3.3911.

[5] M. I. K. Sinapoy, Y. Sibaroni, and S. S. Prasetyowati, "Comparison of LSTM and IndoBERT Method in Identifying Hoax on Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 3, pp. 657–662, Jun. 2023, doi: 10.29207/resti.v7i3.4830.

[6] B. Wibowo, A. Fathl Jannah, and L. Hafiz, "Optimalisasi Bot Telegram untuk Deteksi Situs Perjudian Online di Dunia Pendidikan dan Sektor Pemerintah," *Jurnal Pengabdian Masyarakat Sultan Indonesia*, vol. 2, no. 1, pp. 17–24, Dec. 2024, doi: 10.58291/abdisultan.v2i1.316.

[7] K. Setyo Nugroho, I. Akbar, and A. Nizar Suksmawati, "Deteksi Depresi Dan Kecemasan Pengguna Twitter Menggunakan Bidirectional LSTM," in *CIASTECH 2021*, Malang: Universitas Widyagama Malang, Dec. 2021, pp. 287–296.

[8] S. Choi, “Understanding Involuntary Illegal Online Gamblers in the U.S.: Framing in Misleading Information by Online Casino Reviews,” *UNLV Gaming Research & Review Journal*, vol. 27, no. 1, pp. 23–47, Apr. 2023, doi: 10.9741/2327-8455.1474.

[9] A. Hernández-Ruiz and Y. Gutiérrez, “Analysing the Twitter accounts of licensed Sports gambling operators in Spain: a space for responsible gambling?,” *Communication & Society*, vol. 34, no. 4, pp. 65–79, Oct. 2021, doi: 10.15581/003.34.4.65-79.

[10] P. Angellina and B. Prasetyo, “Pertanggungjawaban Pidana Terhadap Pelaku yang Mempromosikan Judi Online,” *Ranah Research : Journal of Multidisciplinary Research and Development*, vol. 7, no. 2, pp. 946–952, Dec. 2024, doi: 10.38035/rrj.v7i2.1395.

[11] K. Kolandai-Matchett and M. Wenden Abbott, “Gaming-Gambling Convergence: Trends, Emerging Risks, and Legislative Responses,” *Int J Ment Health Addict*, vol. 20, no. 4, pp. 2024–2056, Aug. 2022, doi: 10.1007/s11469-021-00498-y.

[12] A. Bradley and R. J. E. James, “How are major gambling brands using Twitter?,” *Int Gambl Stud*, vol. 19, no. 3, pp. 451–470, Sep. 2019, doi: 10.1080/14459795.2019.1606927.

[13] T. Teichert, A. Graf, T. B. Swanton, and S. M. Gainsbury, “The joint influence of regulatory and social cues on consumer choice of gambling websites: preliminary evidence from a discrete choice experiment,” *Int Gambl Stud*, vol. 21, no. 3, pp. 480–497, Sep. 2021, doi: 10.1080/14459795.2021.1921011.

[14] S. Tang, X. Mi, Y. Li, X. Wang, and K. Chen, “Clues in Tweets: Twitter-Guided Discovery and Analysis of SMS Spam,” in *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, New York, NY, USA: ACM, Nov. 2022, pp. 2751–2764. doi: 10.1145/3548606.3559351.

[15] E. Zhu, J. Wu, H. Liu, and K. Li, “A Sentiment Index of the Housing Market in China: Text Mining of Narratives on Social Media,” *The Journal of Real Estate Finance and Economics*, vol. 66, no. 1, pp. 77–118, Jan. 2023, doi: 10.1007/s11146-022-09900-5.

[16] P. Chiawchansilp and P. Kantavat, “Spam Article Detection on Social Media Platform Using Deep Learning: Enhancing Content Integrity and User Experience,” in *Proceedings of the 13th International Conference on Advances in Information Technology*, in IAIT ’23. New York, NY, USA: Association for Computing Machinery, 2023. doi: 10.1145/3628454.3628459.

[17] S. Kaddoura, S. A. Alex, M. Itani, S. Henno, A. AlNashash, and D. J. Hemanth, “Arabic spam tweets classification using deep learning,” *Neural Comput Appl*, vol. 35, no. 23, pp. 17233–17246, 2023, doi: 10.1007/s00521-023-08614-w.

[18] M. Liu, Y. Zhang, B. Liu, Z. Li, H. Duan, and D. Sun, “Detecting and Characterizing SMS Spearphishing Attacks,” in *Proceedings of the 37th Annual Computer Security Applications Conference*, in ACSAC ’21. New York, NY, USA: Association for Computing Machinery, 2021, pp. 930–943. doi: 10.1145/3485832.3488012.

[19] N. Nasir, F. Iqbal, M. Zaheer, M. Shahjahan, and M. Javed, “Lures for Money: A First Look into YouTube Videos Promoting Money-Making Apps,” in *Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security*, in ASIA CCS ’22. New York, NY, USA: Association for Computing Machinery, Mar. 2022, pp. 1195–1206. doi: 10.1145/3488932.3517404.