INFORMATICS AND DIGITAL EXPERT (INDEX) - VOL. 7 NO. 2 (2025) 98-104
Terbit *online* pada laman web jurnal : https://e-journal.unper.ac.id/index.php/informatics
| ISSN (Print) 2775-2208 | ISSN (Online) 2715-0453 |

Analisis Komparatif Algoritma *One-Class* Untuk Klasifikasi Teks Iklan Judi *Online* Berbasis *Embedding* IndoBERT

Muhamad Randi Septiansah[1], Cepy Slamet[2], Gitarja Sandi[3]

[123]*UIN Sunan Gunung Djati Bandung, Jl. AH. Nasution No.105, Cibiru, Kota Bandung, 40614, Indonesia*
e-mail: 1217050086@student.uinsgd.ac.id[1], cepy_lucky@uinsgd.ac.id[2], sandi@uinsgd.ac.id[3]

INFORMASI ARTIKEL

Sejarah Artikel:
Diterima Redaksi : 12 Juli 2025
Revisi Akhir : 05 November 2025
Diterbitkan *Online* : 10 November 2025

Kata Kunci :
Deteksi Anomali, IndoBERT, *One-Class* SVM, *Autoencoder*, Iklan Judi *Online*

Korespondensi :
Telepon / Hp : +62 85155026211
E-mail :
1217050086@student.uinsgd.ac.id

A B S T R A K

Penyebaran iklan judi *online* di Indonesia terus meningkat meskipun berbagai upaya pemblokiran telah dilakukan. Iklan-iklan ini menyebar luas melalui media sosial dan situs daring, sehingga dibutuhkan sistem otomatis yang mampu menyaring konten iklan judi secara efektif. Penelitian ini mengusulkan pendekatan deteksi anomali satu kelas (*one-class anomaly detection*) sebagai metode penyaringan awal berbasis pembelajaran hanya dari data teks judi. Untuk merepresentasikan teks, digunakan *contextual embedding* IndoBERT yang menghasilkan vektor fitur berdimensi 768. Selanjutnya, dilakukan analisis komparatif terhadap tiga algoritma deteksi anomali: *One-Class* SVM, *Isolation Forest*, dan *Autoencoder*. Pengujian dilakukan dalam dua skenario: data seimbang dan tidak seimbang. Pada skenario seimbang, OCSVM menunjukkan performa terbaik dengan *F1-score* mencapai 91%. Sementara itu, dalam skenario tidak seimbang yang meniru kondisi dunia nyata, model *Autoencoder* mendapatkan *recall* terbaik mencapai 94% dengan jumlah *false positive* yang sangat rendah. Hasil ini menunjukkan bahwa *Autoencoder* layak direkomendasikan sebagai sistem penyaringan awal sebelum proses klasifikasi lanjutan atau validasi manual dilakukan.

1. PENDAHULUAN

Meskipun upaya pemblokiran oleh pemerintah terus dilakukan, aksesibilitas situs judi *online* bagi pengguna internet di Indonesia faktanya masih sangat tinggi, salah satunya akibat masifnya penyebaran informasi melalui iklan [1].

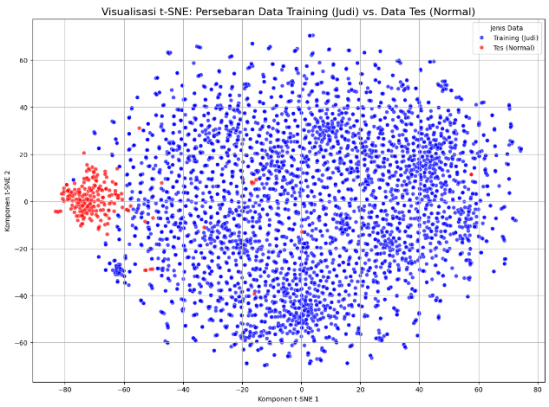
Pembatasan akses terhadap iklan judi *online* ini menjadi strategi krusial untuk menekan pertumbuhan pemain baru, sejalan dengan amanat Undang-Undang ITE Pasal 27 Ayat 2 yang melarang konten perjudian [2]. Urgensi masalah ini dipertegas oleh data dari Pusat Pelaporan dan Analisis Transaksi Keuangan (PPATK) yang mencatat perputaran uang judi *online* diproyeksikan mencapai Rp1.200 triliun dengan 8,8 juta pemain hingga tahun 2025, di mana promosi melalui iklan di berbagai media menjadi salah satu pendorong utamanya [3].

Dampak langsung dari paparan iklan ini terkonfirmasi melalui survei yang dilakukan oleh Populix pada tahun 2023. Studi bertajuk "*Understanding the Impact of Online Gambling Ads Exposure*" tersebut menemukan bahwa 84% dari 1.058 responden telah melihat iklan judi *online* dalam enam bulan terakhir. Lebih lanjut, survei tersebut mengungkapkan bahwa 16% responden mengaku langsung mencoba bermain judi *online* setelah terpapar oleh iklan tersebut di media sosial atau situs web [4].

Data ini mengindikasikan bahwa iklan judi *online* memiliki pengaruh yang signifikan sebagai pemicu konversi pengguna internet menjadi pemain baru, menyoroti pentingnya pengembangan sistem deteksi iklan otomatis yang efektif

Salah satu tantangan utama dalam membangun sistem deteksi iklan judi adalah keragaman bentuk dan topik teks normal yang sangat luas, terutama di media sosial. Mendeteksi promosi judi yang proporsinya sangat kecil di antara lautan teks umum merupakan tantangan teknis tersendiri yang membutuhkan pendekatan yang adaptif dan selektif [5]. Keragaman topik, gaya bahasa, dan kosakata dalam kategori teks 'non-judi' atau teks normal sangatlah luas dan hampir tak terbatas, sehingga sulit untuk direpresentasikan secara menyeluruh dalam satu dataset.

Oleh karena itu, penelitian ini mengusulkan pendekatan alternatif yang lebih praktis, yaitu deteksi anomali satu kelas (*one-class anomaly detection*) [6]. Pendekatan ini melatih model hanya pada contoh-contoh teks judi untuk mempelajari pola intrinsiknya, dengan tujuan agar model mampu mengidentifikasi teks serupa dan menganggap semua jenis teks lainnya sebagai data normal (*outlier*) [7], [8] sebagaimana di visualisasikan pada gambar 1.



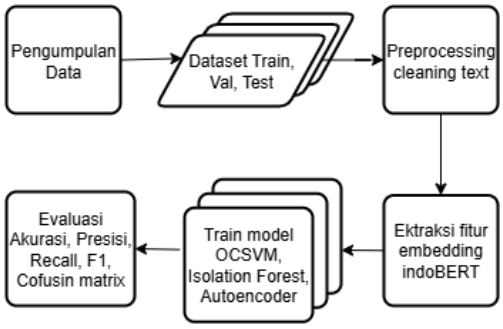
Gambar 1. Visualisasi persebaran dataset teks judi dan teks normal yang telah di ektraksi oleh *embedding* IndoBERT.

Untuk merepresentasikan teks secara efektif, penelitian ini memanfaatkan *contextual embedding* dari model Transformer IndoBERT (*indobenchmark /indobert-base-p1*) [9], [10]. Representasi vektor 768 dimensi ini digunakan sebagai input seragam untuk mengevaluasi dan membandingkan tiga paradigma algoritma deteksi anomali yang berbeda. Ketiga model tersebut adalah *One-Class* SVM yang berbasis *support* vector [11], *Isolation Forest* yang berbasis *ensemble tree* [12], dan *Autoencoder* yang berbasis *deep learning* [13].

Kontribusi utama dari penelitian ini adalah analisis komparatif terhadap tiga arsitektur *one-class* untuk menentukan model yang paling andal dan efektif sebagai sistem penyaringan awal konten iklan judi *online*. Performa setiap model tidak hanya dievaluasi pada dataset tes yang seimbang, tetapi juga diuji dalam kondisi tidak seimbang yang merefleksikan skenario dunia nyata. Analisis ini bertujuan untuk memberikan rekomendasi model yang tidak hanya akurat secara teknis, tetapi juga paling layak diimplementasikan sebagai filter otomatis tahap awal, sebelum proses klasifikasi lanjutan atau validasi manual dilakukan

2. METODE PENELITIAN

Alur metode penelitian dapat dilihat pada Gambar 2.



Gambar 2. Alur Metode Penelitian

Alur sistem dimulai dari tahap pengumpulan dan pembagian data menjadi *set training* dan *testing*. Data *training* yang telah melalui tahap pra-pemrosesan untuk pembersihan dan standardisasi teks kemudian diekstraksi fiturnya menggunakan metode *contextual embedding* IndoBERT. Vektor fitur yang dihasilkan kemudian digunakan sebagai input yang seragam untuk melatih tiga model deteksi anomali yang berbeda secara komparatif: *One-Class* SVM, *Isolation Forest*, dan *Autoencoder*. Tahap akhir adalah pengujian model, di mana setiap model yang telah dilatih dievaluasi performanya menggunakan data testing yang telah disisihkan sebelumnya untuk menghasilkan metrik evaluasi akhir sebagai kesimpulan penelitian.

2.1. Pengumpulan Data

Data teks iklan judi *online* dikumpulkan dari berbagai platform media sosial dan situs daring. Beberapa sumber yang digunakan antara lain adalah Instagram, Facebook, Telegram, Twitter (X), serta beberapa situs *streaming* seperti Anoboy, Indian River Bar, dan Maroon Bells Morris. Data dikumpulkan dengan dua metode utama: pertama, melalui penyalinan langsung teks dari *caption* atau deskripsi pada postingan media sosial; dan kedua, melalui proses ekstraksi teks dari gambar menggunakan teknologi *Optical Character Recognition* (OCR). OCR digunakan untuk mengekstrak informasi yang terkandung dalam konten visual, seperti poster atau tangkapan layar iklan.

Untuk keperluan evaluasi model, dua set data normal (non-judi) yang berbeda telah disiapkan untuk mengakomodasi dua skenario pengujian. Pada skenario pengujian seimbang, sampel teks normal diambil dari dataset publik dataset_sms_spam_v2 (label 'ham') yang bersumber dari repositori GitHub. Sementara itu, untuk skenario pengujian tidak seimbang yang dirancang untuk menyimulasikan kondisi dunia nyata, kumpulan teks normal dihimpun dari berbagai sumber daring yang lebih beragam. Sumber-sumber ini mencakup judul berita, kolom komentar media sosial, unggahan Twitter, ulasan aplikasi, dan ulasan produk untuk menciptakan set data yang lebih menantang dan realistis.

2.2. *Preprocessing* Data

Sebelum ekstraksi fitur dilakukan, tahap pra-pemrosesan data dilakukan untuk membersihkan dan menyeragamkan seluruh teks dalam *dataset*. Untuk menjaga konsistensi, proses ini dimulai dengan membagi *dataset* untuk *train*, evaluasi, dan *test* supaya data evaluasi dan *test* tetap terjaga dari kebocoran pada tahap augmentasi dan pelatihan model.

Selanjutnya mengecilkan *case*, di mana seluruh teks diubah menjadi format huruf kecil. Untuk menghilangkan elemen yang tidak relevan yang dapat mengganggu proses pembelajaran model, seperti URL, mention pengguna, angka, dan semua karakter kecuali spasi dan huruf, tahap penghapusan *emoticon* juga dilakukan.

Metode *augmentasi* data digunakan untuk memperbesar dan memperkaya jumlah data pelatihan karena jumlah data asli yang digunakan untuk pelatihan sangat terbatas. Dengan menggunakan *Large Language Model* (LLM) Gemini 2.5 pro, empat variasi teks baru dibuat berdasarkan setiap sampel dari data pelatihan asli. Untuk menghasilkan data sintetis yang beragam, *augmentasi* ini dirancang secara kualitatif dengan menargetkan empat gaya bahasa tertentu: sopan, meyakinkan, santai, dan agresif.

2.3. Ektraksi Fitur

Dalam penelitian ini, model *pre-trained indobenchmark/indobert-base-p1* digunakan. Setiap teks yang sudah bersih dimasukkan ke dalam model, dan satu vektor fitur tunggal dengan 768 dimensi diekstraksi untuk merepresentasikan keseluruhan teks. Vektor ini didapatkan melalui strategi *mean-pooling* dengan menghitung nilai rata-rata dari seluruh *output* token pada lapisan tersembunyi terakhir (*last hidden state*) dari model.

2.4. Pelatihan Model

Eksperimen ini berfokus pada tahap pelatihan model, di mana tiga model deteksi anomali *One-Class* SVM, *Isolation Forest*, dan *Autoencoder* diuji kemampuannya untuk menemukan teks judi *online*. Untuk memastikan perbandingan yang komprehensif, setiap model dilatih dan dievaluasi secara independen menggunakan representasi fitur yang seragam, yaitu vektor berdimensi 768 dari IndoBERT. Sesuai dengan pendekatan *one-class*, proses pelatihan untuk setiap model dilakukan secara eksklusif pada data training yang berisi sampel kelas judi.

Konfigurasi *hyperparameter* terbaik untuk setiap model, yang diperoleh dari proses *tuning* pada data validasi, adalah sebagai berikut.

- *One-Class* SVM
  kernel='rbf'
  gamma=’scale’
  nu=0.001

- *Isolation Forest*
  n_estimators=100
  contamination=0.03.

- *Autoencoder*
  encoding_dim=32
  loss=’mean_squared_error’
  best_threshold=0.018556

2.5. Evaluasi

Setelah melatih model, model perlu dinilai kinerja klasifikasinya, adapun 4 metrik utama yaitu *Accuracy*, *Precision*, *recall*, dan *F1-score* yang didasarkan pada kompenen *confusion matrix* [14] sebagai berikut.

Tabel 1. Komponen *Confusion matrix*

	Prediksi Positif	**Prediksi Negatif**
Nyata Positif	*True Positive* (TP)	*False Negative* (FN)
Nyata Negatif	*False Positive* (FP)	*True Negative* (TN)

1. *Accuracy*
   Proporsi prediksi yang benar dari seluruh sampel.
   $$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$
2. *Precision*
   Proporsi prediksi positif yang benar (mengukur keandalan prediksi positif).
   $$Precision = \frac{TP}{TP + FP} \quad (2)$$
3. *Recall*
   Proporsi kasus positif yang berhasil terdeteksi.
   $$Recall = \frac{TP}{TP + FN} \quad (3)$$
4. *F1-score*
   Rata rata harmonik antara *precision* dan *recall*.
   $$F1 = 2\ x\ \frac{Precision\ x\ Recall}{Precision + Recall} \quad (4)$$

Evaluasi performa model dalam penelitian ini dilakukan melalui dua skenario pengujian yang berbeda untuk mengukur aspek performa yang berlainan. Skenario pertama adalah pengujian pada set data seimbang, yang bertujuan untuk membandingkan kapabilitas dasar setiap algoritma secara adil dalam membedakan kelas positif dan negatif. Pada skenario ini, performa utama diukur menggunakan metrik *F1-Score* yang menyeimbangkan *precision* dan *recall*.

Skenario kedua adalah pengujian pada set data tidak seimbang, yang dirancang untuk menyimulasikan kondisi implementasi di dunia nyata. Tujuan utamanya adalah untuk mengukur ketangguhan model terhadap "alarm palsu" (*false positive*), sehingga metrik yang menjadi fokus utama adalah *Recall*. Hasil dari kedua skenario ini kemudian dianalisis secara komparatif untuk memberikan kesimpulan yang komprehensif mengenai model mana yang paling efektif dan praktis.

2.6. Lingkungan Eksperimen

Seluruh eksperimen dalam penelitian ini dijalankan pada platform *cloud computing* Google Colaboratory. Lingkungan komputasi ini didukung oleh CPU Intel(R) Xeon(R), RAM sistem sebesar 12.7 GB, dan diakselerasi oleh NVIDIA T4 GPU dengan VRAM 16 GB untuk menangani proses komputasi yang intensif. Dari sisi perangkat lunak, eksperimen diimplementasikan menggunakan bahasa pemrograman Python (versi 3.10) pada sistem operasi berbasis Linux. *Library* utama yang dimanfaatkan meliputi Pandas untuk manipulasi data, NumPy untuk operasi numerik, Scikit-learn untuk implementasi model OCSVM dan *Isolation Forest* serta metrik evaluasi, TensorFlow untuk membangun model Autoencoder, dan *library Transformers* untuk memuat model IndoBERT.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Data teks judi *online* di kumpulkan dari beberapa media sosial yaitu Instagram sebanyak 1.095, Facebook sebanyak 9 baris data, Telegram sebanyak 12 baris data, Twitter atau X sebanyak 53 baris data, dari situs *streaming* (*anoboy, indianriverbar, maroonbellsmorris)* sebanyak 28 baris data dengan total 1.250 dataset yang terkumpul. Data normal di dapat dari github agtbaskara/dataset_sms_spam_v2 hanya di ambil data *ham* sebanyak 200 untuk menjadi data *test* data seimbang di bagian evaluasi.

Untuk pengujian tidak seimbang didapat dari banyak sumber yaitu judul berita, kolom komentar media sosial, unggahan Twitter, ulasan aplikasi, dan ulasan produk dengan total 198.941 baris data teks normal.

Tabel 2. Teks Judi dan Teks Normal

Teks Judi	Teks Normal
SCATTER HITAM AUTO SULTAN EXCLUSIVE MAHJONG WINS 3 BLACK SCATTER VIP SCATTER LIGA JUARA GAMPANG MAXWIN	Maaf jika ada janji yang belum terpenuhi, jika ada janji boleh mengingatkan saya.

3.2. Preprocessing Data

Sebelum dilakukan ekstraksi fitur, seluruh dataset telah melalui serangkaian tahap pra-pemrosesan untuk memastikan kualitas dan konsistensi data. Sebagai langkah krusial untuk menjaga validitas evaluasi, 1.250 data asli teks judi (kelas positif) telah dibagi menjadi tiga set: 850 data untuk *training*, 200 data untuk *validation*, dan 200 data untuk *testing*. Untuk mengakomodasi dua skenario evaluasi, data tes kelas positif tersebut kemudian digunakan secara berbeda. Pada skenario pengujian seimbang, 200 data judi digabungkan dengan 200 data teks normal dari dataset publik. Sedangkan pada skenario pengujian tidak seimbang, 200 data judi yang sama digabungkan dengan 198.941 data teks normal yang telah dihimpun dari berbagai sumber daring untuk menyimulasikan kondisi dunia nyata.

Data *training* telah dilakukan *augmentasi* untuk menambah jumlah data *training* dengan empat gaya *augmentasi* sopan, meyakinkan, santai, agresif. Total data training menjadi 4.250, hasil *augmentasi* data training dapat dilihat pada tabel 3.

Tabel 3. Hasil *Augmentasi*

Gaya *Augmentasi*	Hasil
Data Asli	Kimtoto RTP HARI INI SEGERA MAINKAN GAME GACOR DIKIMTOTO AKURAT DAN SUDAH PASTI JP GOOGLE KIMTOTO DAFTAR SEKARANG
Sopan	Silakan periksa RTP terbaru hari ini di Kimtoto dan mainkan game gacor untuk kesempatan menang yang lebih tinggi.
Meyakinkan	Mainkan game paling akurat berdasarkan RTP hari ini di Kimtoto. Kemenangan JP sudah pasti di depan mata.
Santai	Cek RTP gacor hari ini di Kimtoto, bro. Udah pasti JP, langsung daftar aja.
Agresif	RTP PALING AKURAT HARI INI DI KIMTOTO! MAINKAN GAME GACOR SEKARANG DAN PASTI JP! GOOGLE KIMTOTO!

Proses pembersihan kemudian diterapkan pada semua data, yang meliputi konversi teks menjadi huruf kecil (*case folding*) serta penghapusan elemen non-semantik seperti URL, *mention* pengguna, angka, emoticon, dan karakter spesial lainnya. Tabel Hasil *cleaning* data dapat dilihat pada tabel 4.

Tabel 4. Hasil *Cleaning*

Teks Asli	Teks *Cleaning*
SAKTI788 REKOMENDASI SLOT GACOR BULAN MAY GATES OF OLYMPUS SUPER SCATTER WAKTUNYA BANJIR SCATTER DIBULAN MEY!	sakti788 rekomendasi slot gacor bulan may gates of olympus super scatter waktunya banjir scatter dibulan mey

3.3. Evaluasi Skenario Seimbang

Seluruh data yang telah dipersiapkan pada tahap sebelumnya kemudian direpresentasikan sebagai vektor fitur menggunakan *embedding* IndoBERT berdimensi 768. Vektor-vektor inilah yang digunakan sebagai input seragam untuk melatih dan menguji ketiga model deteksi anomali: *One-Class* SVM, *Isolation Forest*, dan *Autoencoder*.

Untuk mengetahui model dengan performa paling optimal, evaluasi akhir dilakukan pada 400 data tes murni (200 teks judi dan 200 teks normal). Tabel 5 merangkum perbandingan hasil skenario seimbang dari setiap model yang telah dijalankan dengan konfigurasi *hyperparameter* terbaiknya.

Tabel 5. Hasil Evaluasi Skenario Seimbang

Metrik	OCSVM	*Isolation Forest*	Autoencoder
F1-Score	0.91	0.86	0.91
Precision	0.92	0.83	0.90
Recall	0.91	0.84	0.91
Accuracy	0.91	0.85	0.91

Untuk analisis kesalahan yang lebih mendalam, Tabel 6 menyajikan rincian *Confusion Matrix* dari setiap model. Hasil rekapitulasi dari *Confusion Matrix* untuk setiap model yang telah di-*tuning* pada 200 data tes positif (Judi) dan 200 data tes negatif (Normal) disajikan pada Tabel 6. di bawah ini.

Tabel 6. *Confusion Matrix*

Model	TP	TN	FP	FN
OCSVM	181	184	19	16

Model	TP	TN	FP	FN
Isolation Forest	166	173	27	34
Autoencoder	180	183	17	20

Berdasarkan hasil pada skenario seimbang (Tabel 5 dan 6), terlihat bahwa model *One-Class* SVM dan *Autoencoder* menunjukkan performa terbaik dengan *F1-Score* identik sebesar 0.91. OCSVM menunjukkan keunggulan pada *Precision* (0.92), yang berarti ia paling sedikit melakukan kesalahan dalam menuduh teks normal sebagai judi (FP terendah, hanya 16 kasus). Di sisi lain, *Autoencoder* dan OCSVM sama-sama menunjukkan *Recall* tertinggi (0.91), membuktikan kemampuannya dalam menangkap mayoritas teks judi. *Isolation Forest,* meskipun performanya baik, berada sedikit di bawah kedua model lainnya dalam skenario ini.

3.4. Evaluasi Skenario Tidak Seimbang

Setelah mengetahui performa dasar, pengujian dilanjutkan ke skenario tidak seimbang untuk mengukur kelayakan praktis setiap model. Pengujian ini menggunakan set data yang jauh lebih besar dan realistis, terdiri dari 200 teks judi dan 198.941 teks normal. Dalam skenario ini, metrik *Precision* menjadi tolok ukur paling krusial untuk menilai apakah model dapat diimplementasikan tanpa menghasilkan terlalu banyak alarm palsu (*false positive*).

Tabel 7. Metrik Skenario Tidak Seimbang

Metrik	OCSVM	*Isolation Forest*	Autoencoder
F1-Score	0.02	0.02	0.02
Precision	0.01	0.01	0.01
Recall	0.93	0.83	0.94
Accuracy	0.91	0.89	0.91

Karena data tidak seimbang, dibutuhkan analisis kesalahan yang lebih mendalam, Tabel 8 menyajikan *confusion matrix* dari setiap model yang telah di evaluasi.

Tabel 8. *Confusion Matrix* Skenario Tidak Seimbang

Model	TP	TN	FP	FN
OCSVM	185	179470	15	19471
Isolation Forest	166	177434	34	21057
Autoencoder	189	180860	11	18081

Evaluasi pada data tidak seimbang dengan total 198.941 sampel menunjukkan bahwa ketiga model deteksi anomali menghasilkan nilai *recall* yang tinggi, namun memiliki nilai precision yang sangat rendah. Berdasarkan Tabel 7, *Autoencoder* mencatat nilai recall tertinggi sebesar 0,94 dengan false positive (FP) paling sedikit yaitu 11 kasus, sebagaimana tercantum dalam Tabel 8. OCSVM menyusul dengan *recall* sebesar 0,93 dan 15 kasus FP.

Sementara itu, *Isolation Forest* memperoleh *recall* 0,83 dengan jumlah FP tertinggi sebesar 34. Meskipun ketiga model memiliki precision yang sama (0,01), tingkat *accuracy* tetap tinggi di atas 0,89, yang dipengaruhi oleh dominasi data negatif dalam skenario tidak seimbang.

3.5. Analisis Performa Skenario Seimbang

Evaluasi terhadap data seimbang dilakukan guna mengetahui sejauh mana masing-masing model mampu membedakan teks iklan judi secara adil. Berdasarkan Tabel 5, *One-Class* SVM dan *Autoencoder* sama-sama mencatat F1-*Score* tertinggi sebesar 0,91, sedangkan *Isolation Forest* tertinggal dengan skor 0,86.

Dari sisi presisi dan sensitivitas, terlihat bahwa setiap model memiliki kekuatan yang berbeda. OCSVM mencatat presisi paling tinggi (0,92), menunjukkan bahwa model ini mampu meminimalkan kesalahan dalam mengidentifikasi data normal sebagai teks judi, yang tercermin dari jumlah *false positive* yang hanya 19 kasus. Di sisi lain, *Autoencoder* menunjukkan performa paling stabil, dengan nilai *Precision* dan *Recall* yang seimbang di angka 0,90 dan 0,91. Sementara itu, *Isolation Forest* menunjukkan hasil paling rendah dalam kedua metrik tersebut, dengan *Recall* 0,84 dan *Precision* 0,83, serta menghasilkan *false negative* terbanyak, yaitu 34, yang berarti lebih banyak kasus judi terlewatkan.

Secara umum, perpaduan antara kekuatan representasi fitur dari IndoBERT dan kemampuan deteksi model menunjukkan hasil paling optimal pada OCSVM dapat dipertimbangkan sebagai model paling optimal untuk data seimbang, menjadikannya kandidat yang cocok untuk diterapkan dalam sistem penyaringan awal klasifikasi judi *online*.

3.6. Analisis *Erorr* Kualitatif

Untuk memahami kelemahan setiap model, dilakukan analisis pada sampel-sampel yang salah diklasifikasikan. Ditemukan dua pola kesalahan utama.

Pertama, kesalahan *False Positive* (teks normal dianggap judi) umumnya terjadi karena adanya kemiripan konseptual antara teks normal dengan teks promosi. Model mendeteksi konsep abstrak seperti transaksi (kata saldo, PIN), penawaran (paket, harga), dan ajakan bertindak (wajib kumpul, jangan lupa) yang juga umum dalam teks judi. Selain itu, teks dengan struktur tidak umum seperti bahasa teknis (sql, *database*) juga cenderung dianggap sebagai anomali, sama seperti teks judi.

Kedua, kesalahan *False Negative* (teks judi gagal dideteksi) seringkali disebabkan oleh variasi gaya bahasa yang ekstrem yang tidak cukup terwakili dalam data *training*. Ini mencakup teks bergaya "teriakan merek" yang sangat pendek dan menggunakan huruf kapital (PATRIOT77), teks bergaya "kueri pencarian" (LINK SLOT GACOR), dan teks dengan struktur unik seperti jadwal jam gacor, yang *embedding*-nya berada jauh dari pusat distribusi data *training*.

3.7. Analisis Performa Skenario Tidak Seimbang

Berdasarkan hasil evaluasi, skenario tidak seimbang menguji ketahanan model dalam kondisi nyata dengan distribusi data yang sangat timpang. Dalam konteks ini, recall menjadi indikator kunci karena menggambarkan kemampuan model dalam mengenali teks iklan judi di antara dominasi besar teks normal. *Autoencoder* menampilkan performa paling optimal karena mampu mencapai *recall* tertinggi sekaligus menjaga jumlah *false positive* pada tingkat terendah dibandingkan dua model lainnya. Hanya terdapat 11 kasus FP dari hampir 199.000 data negatif, yang menunjukkan bahwa Autoencoder memiliki sensitivitas tinggi tanpa mengorbankan ketelitian dalam membedakan teks anomali dari teks normal.

OCSVM menunjukkan karakteristik performa yang serupa, meskipun sedikit lebih tinggi dari sisi kesalahan klasifikasi negatif, yakni 15 FP. *Isolation Forest,* meski tetap mempertahankan *recall* di atas 0,8, menunjukkan hasil yang lebih fluktuatif, dengan 34 kasus FP dan recall yang lebih rendah dibandingkan dua model lainnya.

Secara keseluruhan, arsitektur *Autoencoder* menunjukkan ketahanan dan konsistensi dalam memproses data skala besar dengan ketidakseimbangan distribusi, menjadikannya model yang paling optimal dalam skenario ini berdasarkan kombinasi antara *recall* tinggi dan tingkat *false positive* yang sangat rendah.

4. KESIMPULAN

Dalam upaya mencari metode deteksi teks judi *online* yang paling efektif, penelitian ini mengevaluasi tiga pendekatan *one-class* berbeda: OCSVM, *Isolation Forest*, dan *Autoencoder*. Evaluasi awal pada data seimbang menunjukkan bahwa OCSVM dan Autoencoder merupakan kandidat paling menjanjikan, dengan *F1-skor* yang sama-sama mencapai 91% dan unggul dibandingkan *Isolation Forest* dalam hal stabilitas metrik. Namun, pengujian pada skenario tidak seimbang yang mencerminkan kondisi dunia nyata menunjukkan bahwa Autoencoder memiliki keunggulan lebih lanjut, dengan *recall* tertinggi sebesar 94% dan jumlah *false positive* yang sangat rendah. Temuan ini menegaskan bahwa *Autoencoder* lebih layak direkomendasikan sebagai sistem penyaringan awal konten iklan judi *online*, karena mampu mempertahankan performa yang andal meskipun dihadapkan pada distribusi data yang tidak seimbang sebuah tantangan umum yang sering menyebabkan penurunan performa pada sistem deteksi anomali.

DAFTAR PUSTAKA

[1] S. H. Bakhtiar and A. N. Adilah, "Fenomena Judi Online : Faktor, Dampak, Pertanggungjawaban Hukum," *Innovative: Journal Of Social Science Research*, vol. 4, no. 3, pp. 1016–1026, May 2024, doi: 10.31004/innovative.v4i3.10547.

[2] PRESIDEN REPUBLIK INDONESIA, "UNDANG UNDANG REPUBLIK INDONESIA NOMOR 1 TAHUN 2024," 2024. Accessed: May 12, 2025. [Online]. Available: https://peraturan.bpk.go.id/Download/332870/UU%20Nomor%201%20Tahun%202024.pdf

[3] PPATK, "Promensisko 2025: Menjawab Ancaman Judi Online dan Kejahatan Digital Lewat Aksi," PPATK. Accessed: May 12, 2025. [Online]. Available: https://www.ppatk.go.id/siaran_pers/read/1474/promensisko-2025-menjawab-ancaman-judi-online-dan-kejahatan-digital-lewat-aksi-.html

[4] Populix, "84% Pengguna Internet Indonesia Terpapar Iklan Judi Online," POPULIX. Accessed: May 12, 2025. [Online]. Available: https://info.populix.co/articles/iklan-judi-online/

[5] R. Bayu Perdana, I. Budi, A. Budi Santoso, A. Ramadiah, and P. Kresna Putra, "Detecting Online Gambling Promotions on Indonesian Twitter Using Text Mining Algorithm," 2024. [Online]. Available: www.ijacsa.thesai.org

[6] D. Challa, "Comprehensive Review of One-Class Classification Approaches for Anomaly Detection," 2024.

[7] P. Perera, P. Oza, and V. M. Patel, "One-Class Classification: A Survey," Jan. 2021, [Online]. Available: http://arxiv.org/abs/2101.03064

[8] L. Strani, M. Cocchi, D. Tanzilli, A. Biancolillo, F. Marini, and R. Vitale, "One class classification (class modelling): State of the art and perspectives," Feb. 01, 2025, *Elsevier B.V.* doi: 10.1016/j.trac.2024.118117.

[9] B. Wilie *et al.*, "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, K.-F. Wong, K. Knight, and H. Wu, Eds., Suzhou, China: Association for Computational Linguistics, Dec. 2020, pp. 843–857. doi: 10.18653/v1/2020.aacl-main.85.

[10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Oct. 2018, [Online]. Available: http://arxiv.org/abs/1810.04805

[11] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the Support of a High-Dimensional Distribution," *Neural Comput*, vol. 13, no. 7, pp. 1443–1471, Jul. 2001, doi: 10.1162/089976601750264965.

[12] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," in *2008 Eighth IEEE International Conference on Data Mining*, 2008, pp. 413–422. doi: 10.1109/ICDM.2008.17.

[13] J. An and S. Cho, "Variational Autoencoder based Anomaly Detection using Reconstruction Probability," 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:36663713

[14] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*, 3rd ed. 2025. [Online]. Available: https://web.stanford.edu/~jurafsky/slp3/