

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perjudian online tumbuh pesat di Indonesia dalam beberapa tahun terakhir dan menjangkau kelompok pengguna yang sangat luas. Kementerian Komunikasi dan Digital (Komdigi) menindak sekitar 3,7 juta situs dan konten bermuatan judi online sepanjang 20 Oktober 2024 hingga 12 Juli 2026 (Kementerian Komunikasi dan Digital, 2026b). Skala uangnya besar: menurut catatan Pusat Pelaporan dan Analisis Transaksi Keuangan (PPATK), perputaran dana judi online sepanjang 2017 hingga triwulan III 2025 menembus Rp1032 triliun, dengan perputaran tahun 2025 saja mencapai Rp286,84 triliun, turun sekitar 20 persen dari Rp359,81 triliun pada 2024 (PPATK, 2026). Jumlah pemainnya juga melonjak, dari sekitar 3,7 juta orang pada 2023 menjadi 8,8 juta pada 2024, dan 3,8 juta di antaranya diketahui berutang untuk bermain (PPATK, 2025). Paparan bahkan menyentuh anak-anak: hampir 200 ribu anak diperkirakan terpapar, sekitar 80 ribu di antaranya berusia di bawah sepuluh tahun (Kementerian Komunikasi dan Digital, 2026a).

Angka-angka tersebut layak dibaca pelan-pelan. Perputaran ratusan triliun rupiah per tahun berarti dana yang keluar dari ekonomi rumah tangga menuju operator di luar negeri, hampir tanpa jejak pajak. Lonjakan pemain dari 3,7 juta menjadi 8,8 juta hanya dalam setahun menunjukkan hambatan masuknya hampir nol: cukup telepon genggam dan sinyal. Selanjutnya, hampir separuh pemain yang tercatat berutang untuk bermain menandakan persoalan ini sudah melewati tahap hiburan yang kebablasan dan berubah menjadi jerat finansial. Penurunan perputaran pada 2025 memang menunjukkan penindakan mulai berdampak, tetapi angkanya yang masih Rp286,84 triliun menegaskan bahwa upaya yang ada belum cukup.

Dampak judi online tidak berhenti pada kerugian uang. Beberapa penelitian mendokumentasikan pengaruhnya terhadap kondisi ekonomi rumah tangga, tekanan sosial, serta gangguan psikologis seperti kecemasan dan kecanduan (Sriyana, 2025;

Syakira dkk., 2024). Karena banyak korban berasal dari kelompok usia produktif, pelajar, bahkan berpenghasilan rendah, persoalan ini menyangkut kesehatan finansial dan mental sebagian besar masyarakat, bukan sekadar pelanggaran hukum.

Beban itu jarang berhenti pada pemain. Karena uang taruhan umumnya diambil dari penghasilan rumah tangga, kerugian pemain menjadi kerugian keluarga: kebutuhan pokok tergeser, tabungan terkuras, dan sebagian pemain berutang ke pinjaman daring untuk terus bermain (PPATK, 2025). Tekanan itu menjalar menjadi konflik keluarga dan gangguan kesehatan mental, dari kecemasan hingga kecanduan yang sulit diputus tanpa bantuan (Sriyana, 2025; Syakira dkk., 2024). Pada anak dan remaja, paparan iklan judi yang menyaru sebagai permainan membuat batas antara bermain gim dan bertaruh menjadi kabur, sehingga pencegahan sedini mungkin di titik akses, yaitu peramban, menjadi masuk akal.

Kesenjangan literasi digital di dalam keluarga memperparah keadaan. Anggota keluarga yang paling rentan sering kali merupakan pihak yang paling jarang diawasi penggunaannya, sementara orang tua tidak selalu punya kecakapan memasang penyaring atau memeriksa riwayat penjelajahan. Alat perlindungan yang efektif untuk situasi ini harus bekerja otomatis setelah sekali pasang, tidak menuntut pemeliharaan rutin, dan tidak mudah dilewati hanya karena situs berganti alamat. Kebutuhan itulah yang menuntun bentuk akhir solusi pada penelitian ini.

Upaya pemerintah selama ini bertumpu pada pemblokiran alamat situs. Pendekatan ini menekan penyebaran situs judi untuk sesaat, tetapi cepat usang. Yang diblokir umumnya hanya nama domain, yaitu lapisan paling murah dan paling mudah diganti, sementara server pengelolanya berada di luar negeri. Operator hanya butuh beberapa jam untuk menyalin ulang sistem yang sama ke domain baru, sehingga setiap satu situs ditutup, situs pengganti segera bermunculan (Julian dkk., 2025). Akibatnya daftar blokir bersifat reaktif dan selalu tertinggal dari pertumbuhan situs baru. Ketergantungan pada alamat membuat sistem penyaring mudah dilewati.

Praktik berpindah domain ini berjalan sangat sistematis. Operator menyiapkan puluhan domain cadangan (*mirror*) berisi salinan situs yang sama, lalu mengumumkan alamat terbaru lewat iklan media sosial, tautan siaran langsung, dan pesan berantai. Nama domain dibuat mudah ditebak pemainnya, misalnya menambahkan angka di belakang

merek yang sudah dikenal, sehingga pemain lama cepat menemukan alamat baru justru ketika alamat lama diblokir. Bagi sistem penyaring berbasis daftar, setiap pergantian ini berarti satu entri baru yang harus ditemukan, diverifikasi, dan ditambahkan secara manual, sementara bagi operator biayanya hampir nol.

Selain berpindah alamat, operator juga menumpang pada infrastruktur yang sah. Subdomain pada layanan *hosting* gratis, situs kampus atau lembaga yang diretas, dan platform blog publik dipakai untuk menitipkan halaman promosi agar lolos dari pemblokiran domain utama. Salah satu contohnya ditemukan sendiri pada data penelitian ini: halaman promosi judi berbahasa Indonesia yang menumpang pada domain dinas pemerintah daerah berakhiran *.org*, sebagaimana dilaporkan pada subbab 4.1.3. Halaman seperti ini hampir mustahil terjangkau daftar blokir berbasis domain karena domain induknya justru sah.

Alternatif yang lebih tahan terhadap pergantian domain adalah menilai isi teks halaman, bukan alamatnya. Selama sebuah halaman memuat pola bahasa khas promosi judi, misalnya ajakan bermain slot, klaim kemenangan, dan istilah setoran, halaman itu tetap dapat dikenali walaupun domainnya baru. Sejumlah penelitian terbaru di Indonesia menunjukkan deteksi berbasis teks bekerja baik untuk konten judi, tetapi sebagian besar berfokus pada komentar atau iklan media sosial, sebagian memakai model besar seperti IndoBERT (Alfiansyah & Muzakir, 2025; Nugraha dkk., 2026) dan sebagian lagi memakai text mining (Perdana dkk., 2024). Deteksi yang langsung menilai isi halaman situs, berjalan secara real-time di peramban, dan ringan untuk dijalankan lokal masih jarang dibahas.

Celah itulah yang hendak diisi penelitian ini. Kebutuhan praktisnya nyata: perlindungan yang bekerja saat pengguna membuka halaman, bukan setelah daftar blokir diperbarui; berjalan pada komputer biasa tanpa server besar; dan keputusannya dapat diperiksa, tidak diberikan tanpa alasan. Ketiganya menuntun pada pilihan teknis yang dijelaskan berikutnya, yaitu model deteksi teks yang ringan dengan mekanisme yang mampu menunjukkan kata pemicu keputusannya.

Penelitian ini mengembangkan sebuah ekstensi Google Chrome yang dipasangkan dengan model klasifikasi teks pada *backend* lokal. Ketika pengguna membuka halaman, ekstensi mengambil teksnya lalu mengirimkannya ke model; bila skor judi melewati

ambang tertentu, halaman diblokir sebelum pengguna berinteraksi lebih jauh. Model dibangun dengan Bidirectional Long Short-Term Memory (BiLSTM) dan mekanisme Bahdanau Attention. BiLSTM merupakan pengembangan dua arah dari LSTM, yaitu jaringan saraf untuk data berurutan yang dirancang mengatasi keterbatasan memori jangka panjang (Hochreiter & Schmidhuber, 1997), sehingga makna tiap kata dipahami dari konteks sebelum dan sesudahnya. Bahdanau Attention menyoroti kata yang paling menentukan keputusan sekaligus menjadi dasar penjelasan keputusan (Bahdanau dkk., 2015). Pendekatan pembelajaran mendalam untuk klasifikasi teks juga digunakan peneliti di Institut Teknologi PLN, misalnya pada klasifikasi pos pengeluaran perjalanan dinas dengan model sekuensial (Susanty, 2023), sedangkan penerapan Convolutional Neural Network untuk tugas deteksi pada ranah lain ditunjukkan Haris dkk. (2023).

Model dibangun di atas dataset 3646 halaman berbahasa Indonesia yang seimbang antara kelas judi dan kelas aman, yang kemudian dibagi menjadi data latih, data validasi, dan data uji. Data dikumpulkan dengan menelusuri (*crawling*) isi halaman dari daftar domain, lalu dibersihkan dan diproses melalui tahapan *Text Processing* sebelum dilatih. Dengan menilai isi teks dan menampilkan kata pemicu sebagai bukti, sistem diharapkan mendeteksi situs judi secara lebih andal dan keputusannya tetap dapat dipertanggungjawabkan, termasuk ketika situs berpindah domain.

Cara kerja sistem secara ringkas sebagai berikut. Saat pengguna membuka sebuah halaman, ekstensi mengambil teks halaman itu dan mengirimkannya ke *backend* lokal. *Backend* menjalankan praproses teks yang persis sama dengan praproses saat pelatihan, lalu model menghasilkan skor probabilitas judi antara 0 dan 1. Skor 0,50 ke atas membuat halaman diganti halaman blokir yang menampilkan kata-kata pemicu sebagai bukti; skor 0,35 sampai kurang dari 0,50 memunculkan peringatan tanpa pemblokiran; skor di bawah 0,35 membiarkan pengguna melanjutkan. Seluruh proses selesai di bawah satu detik di komputer pengguna sendiri, tanpa mengirim data ke pihak ketiga.

Kebaharuan yang ditawarkan terletak pada gabungan beberapa hal yang belum banyak dikerjakan sekaligus: pemodelan urutan kata secara dua arah pada isi teks halaman situs, bukan pada komentar atau iklan media sosial; penggunaan BiLSTM dengan Bahdanau Attention yang ringan sehingga dapat berjalan lokal secara real-time di

peramban; serta penyediaan kata pemicu sebagai bukti keputusan. Karena keputusan didasarkan pada isi teks, sistem tetap mengenali situs judi meskipun domainnya berganti.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, penelitian ini merumuskan pertanyaan berikut.

1. Bagaimana merancang skrip Python untuk mengumpulkan (*crawling*) isi teks halaman situs dari daftar domain judi dan aman yang telah dikumpulkan?
2. Bagaimana melakukan praproses (*Text Processing*) terhadap data teks hasil *crawling* agar siap digunakan untuk melatih model?
3. Bagaimana merancang model LSTM (BiLSTM dengan Bahdanau Attention) untuk mendeteksi dan memblokir situs judi online berbahasa Indonesia secara real-time melalui ekstensi peramban?
4. Bagaimana kinerja model diukur menggunakan *accuracy*, *precision*, *recall*, F1-score, dan ROC-AUC?

1.3 Tujuan Penelitian

Tujuan penelitian ini adalah:

1. merancang skrip Python untuk mengumpulkan (*crawling*) isi teks halaman situs dari daftar domain judi dan aman;
2. melakukan praproses (*Text Processing*) terhadap data teks hasil *crawling* hingga menjadi data terstandar yang siap dilatih;
3. merancang dan membangun model LSTM (BiLSTM dengan Bahdanau Attention) serta mengintegrasikannya ke ekstensi peramban untuk mendeteksi dan memblokir situs judi online secara real-time;
4. mengukur kinerja model menggunakan *accuracy*, *precision*, *recall*, F1-score, dan ROC-AUC pada data uji.

1.4 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut.

1. Penelitian ini memberikan manfaat teoretis dengan mendokumentasikan alur lengkap deteksi konten judi berbahasa Indonesia, mulai dari pengumpulan data lewat

crawling, praproses teks, sampai pemodelan BiLSTM dua lapis dengan Bahdanau Attention. Dokumentasinya mencakup hal yang jarang dirinci pada penelitian sejenis, yaitu cara membangun *dataset* konten judi dari blocklist publik, jejak pembersihan data tahap demi tahap beserta jumlah barisnya, serta perhitungan manual sel LSTM dan *attention* memakai bobot model terlatih.

2. Penelitian ini memberikan manfaat metodologis berupa kerangka kerja pembangunan *dataset* dan pelatihan model yang dapat diulang. Seluruh parameter *crawling*, aturan penyaringan, pembagian data, dan konfigurasi pelatihan dicatat beserta *seed* acaknya, sehingga peneliti lain dapat menjalankan ulang rangkaian yang sama pada ranah teks berbahasa Indonesia lain, misalnya konten pinjaman daring ilegal atau penipuan berkedok investasi.

3. Penelitian ini memberikan manfaat praktis berupa model dan ekstensi peramban yang dapat langsung dipakai memblokir halaman judi secara real-time di komputer pengguna. Bagi masyarakat, khususnya orang tua, ekstensi ini menjadi lapisan perlindungan yang bekerja otomatis tanpa menuntut keahlian teknis; bagi pemerintah dan penyedia layanan internet, pendekatan berbasis isi teks dapat melengkapi pemblokiran berbasis alamat yang selama ini dipakai.

4. Penelitian ini memberikan manfaat bagi ketersediaan data penelitian bahasa Indonesia. *Dataset* 3646 halaman berlabel yang dibangun beserta skrip pembangunnya menjadi sumbangan untuk mengatasi kelangkaan dataset berlabel untuk ranah spesifik, sebab konten judi termasuk jenis data yang paling jarang tersedia secara terbuka karena sifatnya.

1.5 Ruang Lingkup Masalah

Penelitian ini dibatasi pada hal-hal berikut.

1. Objek deteksi adalah isi teks halaman situs berbahasa Indonesia. Halaman berbahasa selain Indonesia, termasuk halaman judi berbahasa Inggris, berada di luar jangkauan sistem dan tidak dijamin terdeteksi.
2. Sistem terdiri atas ekstensi Google Chrome (Manifest V3) dan *backend* FastAPI yang berjalan lokal.

3. Data yang digunakan berjumlah 3646 halaman hasil *crawling*, seimbang antara kelas judi dan kelas aman.
4. Penelitian tidak membahas deteksi berbasis gambar, video, atau audio, dan tidak mencakup aspek penindakan hukum.

Pembatasan tersebut ditetapkan dengan sadar. Fokus pada teks berbahasa Indonesia dipilih karena sasaran utama situs judi yang diteliti adalah pengguna Indonesia, dan penyaringan bahasa membuat kualitas data lebih terjaga. Konsekuensi pembatasan ini bersifat langsung: kosakata model dibangun hanya dari teks berbahasa Indonesia, sehingga halaman berbahasa lain sebagian besar terbaca sebagai token tak dikenal dan skornya rendah. Batas tersebut terbukti pada analisis kesalahan subbab 4.1.3, yaitu satu halaman taruhan berbahasa Inggris lolos penilaian karena kosakatanya asing bagi model. Deteksi berbasis gambar, video, atau audio membutuhkan jalur pemodelan yang sama sekali berbeda dan lebih berat, sehingga tidak sepadan dengan sasaran sistem yang harus ringan di komputer pengguna. Adapun *backend* dijalankan lokal, bukan sebagai layanan daring, agar teks halaman yang dianalisis tidak pernah meninggalkan perangkat pengguna.

Batasan lain yang perlu disebut adalah cakupan datanya. Teks diambil dari halaman depan setiap domain pada rentang April sampai Mei 2026, sehingga dataset merekam kondisi situs judi pada kurun waktu tersebut. Kosakata promosi judi bergeser dari waktu ke waktu mengikuti tren permainan, dan karena itu model perlu dilatih ulang secara berkala dengan data segar; kebutuhan ini disinggung kembali sebagai saran pada bab penutup.

1.6 Kebaharuan Penelitian

Kebaharuan penelitian ini terletak pada objek yang dinilai dan pada tempat penilaian itu dikerjakan. Penelitian terdahulu mengenai deteksi judi online berbahasa Indonesia sebagian besar menilai komentar media sosial atau iklan, dan sebagian lain menilai nama domain, sedangkan penelitian ini menilai isi teks halaman situs secara utuh. Pilihan tersebut menjawab kelemahan pemblokiran berbasis daftar alamat yang cepat usang karena pengelola situs judi mudah berganti domain, sementara pola bahasa promosinya cenderung bertahan.

Kebaharuan kedua terletak pada penempatan model. Penilaian dijalankan secara real-time di sisi pengguna melalui ekstensi peramban yang terhubung dengan layanan lokal, bukan secara berkala di sisi peladen, sehingga teks halaman tidak pernah meninggalkan perangkat pengguna. Kebaharuan ketiga terletak pada keterbukaan keputusan: bobot Bahdanau Attention dipakai bukan hanya untuk menaikkan ketepatan klasifikasi, melainkan juga ditampilkan kepada pengguna sebagai kata pemicu, sehingga setiap keputusan pemblokiran disertai bukti yang dapat diperiksa.