

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pesatnya pemanfaatan *website* turut menghadirkan tantangan baru, khususnya terkait penyebaran konten judi *online*. Berdasarkan data Kementerian Komunikasi dan Digital Republik Indonesia (*KOMDIGI*), hingga tahun 2025 sebanyak 2.087.109 konten terkait judi *online* telah diblokir [1]. Sementara itu, Pusat Pelaporan dan Analisis Transaksi Keuangan memperkirakan (*PPATK*) bahwa perputaran dana judi *online* di Indonesia sejak tahun 2017 hingga kuartal III tahun 2025 telah mencapai Rp1.032 triliun dengan lebih dari 259 juta transaksi [2]. Data tersebut menunjukkan bahwa penyebaran judi *online* masih menjadi permasalahan yang memerlukan penanganan lebih lanjut..

Saat ini, judi telah bertransformasi ke dalam beragam format permainan, seperti judi berbasis uang, taruhan olahraga, kasino virtual, poker, *slot*, dan judi bola *online* [3]. Selain hadir dalam bentuk permainan, judi *online* juga banyak dipromosikan melalui platform digital dalam bentuk iklan, baik berupa teks, gambar, maupun konten visual lain yang dirancang untuk menarik perhatian pengguna [4]. Jika ditinjau dari aspek visual, iklan judi *online* umumnya memiliki ciri khas berupa tampilan yang mencolok, penggunaan elemen permainan, simbol kemenangan, serta teks promosi yang menawarkan peluang keuntungan kepada pengguna [5]. Keberadaan konten tersebut pada *website* berpotensi dapat diakses oleh berbagai kalangan usia, termasuk anak-anak dan remaja, sehingga dapat menimbulkan dampak negatif secara sosial, finansial, psikologis, maupun akademik [3][6]. Oleh karena itu, diperlukan sebuah sistem identifikasi otomatis yang mampu mendeteksi keberadaan konten judi *online* pada *website* sebagai langkah awal penyaringan konten bagi pengguna.

Seiring dengan berkembangnya teknologi kecerdasan buatan (*Artificial Intelligence*), khususnya *Machine Learning* dan *Deep Learning*. Berbagai penelitian mulai mengeksplorasi metode untuk mendeteksi konten digital ilegal, termasuk konten judi *online*. Salah satu pendekatan yang banyak digunakan adalah

arsitektur *multimodal*, yaitu pendekatan yang memproses lebih dari satu jenis data secara bersamaan, seperti gambar dan teks, sehingga sistem dapat memahami konten secara lebih utuh dibandingkan pendekatan unimodal [5]. Dalam konteks identifikasi *website* judi, pendekatan *multimodal* terbukti mampu memanfaatkan informasi visual dari tangkapan layar *website* dan informasi tekstual hasil *OCR* untuk meningkatkan performa klasifikasi [7].

Beberapa penelitian sebelumnya telah mengusulkan pendekatan *multimodal* untuk mendeteksi iklan judi *online*. Penelitian Fajria menggunakan *PaddleOCR* dan *EasyOCR* untuk ekstraksi teks, *IndoBERT* untuk klasifikasi teks, serta *EfficientNet-B0* dan *ResNet-50* untuk klasifikasi gambar. Hasil terbaik diperoleh dari kombinasi *EfficientNet-B0*, *PaddleOCR*, dan *IndoBERT* dengan akurasi sebesar 96,40% [4]. Selain itu, penelitian Alfiansyah dan Muzakir mengembangkan pendekatan *multimodal ensemble* untuk mendeteksi iklan judi *online* Indonesia dengan menggabungkan *CNN*, *BiLSTM*, *IndoBERT*, dan *AutoML* melalui mekanisme *weighted soft-voting*. Model *ensemble* yang diusulkan memperoleh hasil terbaik dengan *accuracy* 95,42% [8].

Meskipun pendekatan tersebut menunjukkan hasil yang baik, sebagian metode *multimodal* terdahulu masih memiliki kelemahan karena keputusan akhir sangat bergantung pada keberhasilan masing-masing modalitas secara terpisah atau *decision-level fusion*. Pada pendekatan *decision-level fusion*, kegagalan salah satu modalitas, seperti kegagalan *OCR* dalam membaca teks akibat ukuran huruf kecil, kontras rendah, *font* tidak umum, atau kualitas gambar rendah, dapat memengaruhi hasil klasifikasi akhir. Kondisi ini berkaitan dengan permasalahan *missing modality* atau *corrupted modality*, yaitu kondisi ketika salah satu modalitas tidak tersedia, rusak, atau tidak dapat diproses secara optimal sehingga performa model *multimodal* dapat menurun [9],[10].

Untuk mengatasi kelemahan tersebut, penelitian ini mengusulkan sistem deteksi menggunakan pendekatan *feature-level fusion*. Teknik ini dilakukan dengan mengambil representasi fitur dari model visual dan model tekstual, kemudian menggabungkannya sebelum proses klasifikasi akhir. Melalui pendekatan ini, sistem diharapkan dapat mempelajari hubungan antar fitur secara lebih adaptif.

Sehingga ketika hasil *OCR* tidak optimal, model masih dapat mempertimbangkan pola visual dari gambar untuk membantu proses identifikasi konten judi *online* [5][7][9].

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana merancang dan mengimplementasikan model *multimodal* dengan arsitektur *Feature-Level Fusion* menggunakan *DeiT-Small* dan *IndoBERT* untuk mengidentifikasi tangkapan layar *website* yang memuat iklan judi *online*?
2. Bagaimana perbandingan ketangguhan antara arsitektur *Feature-Level Fusion* dan *Decision-Level Fusion* pada skenario normal dan *stress test* ketika teks hasil *OCR* tidak tersedia?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Merancang dan mengimplementasikan model *multimodal* berbasis *Feature-Level Fusion* menggunakan model *DeiT-Small* dan *IndoBERT* untuk mengidentifikasi tangkapan layar *website* yang memuat iklan judi *online*.
2. Menganalisis perbandingan ketangguhan arsitektur *Feature-Level Fusion* dan *Decision-Level Fusion* pada skenario normal dan *stress test* ketika teks hasil *OCR* tidak tersedia.

1.4 Batasan Masalah

Penelitian ini memiliki beberapa batasan agar ruang lingkup studi tetap terarah dan sesuai dengan tujuan utama, yaitu sebagai berikut:

1. Penelitian ini hanya fokus untuk mengidentifikasi iklan judi *online* yang berada di sebuah *website*, tanpa mencoba untuk melakukan penindakan seperti pemblokiran dan juga penghapusan konten.

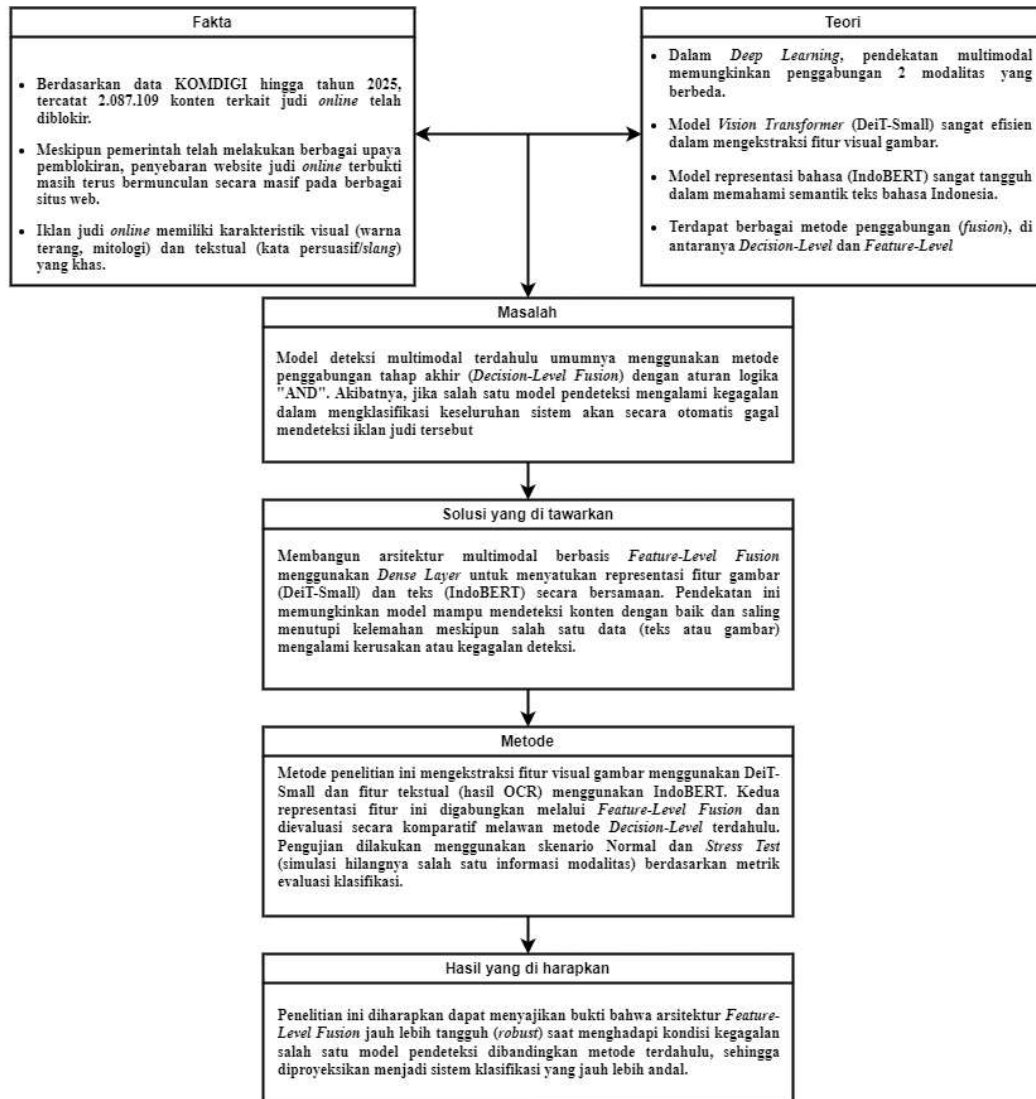
2. Data yang digunakan sebagai *input* awal dari sistem terbatas pada format gambar. Ekstraksi fitur teks didapatkan sepenuhnya menggunakan *Optical Character Recognition (OCR)*.
3. Arsitektur *multimodal* yang dibandingkan dibatasi pada *Feature-Level Fusion* dan *Decision-Level Fusion* dengan aturan logika operator *AND*, menggunakan *DeiT-Small* sebagai pemroses visual dan *IndoBERT* sebagai pemroses teks.
4. Evaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* pada skenario normal serta *missing-text stress test*. Pada stress test, teks hasil *OCR* dikosongkan, sedangkan gambar tangkapan layar tetap tersedia dan tidak diuji dalam kondisi hilang atau rusak.
5. Penelitian mengadaptasi tahapan *CRISP-DM* sampai tahap *evaluation*, sedangkan tahap *deployment* tidak dilakukan karena ruang lingkup penelitian dibatasi pada implementasi model *multimodal* berbasis *Feature-Level Fusion* menggunakan *DeiT-Small* dan *IndoBERT*, serta evaluasi performa model pada skenario normal dan *stress test*.

1.5 Manfaat Penelitian

1. Memberikan kontribusi pada pengembangan *computer vision* dan *NLP* khususnya pada penerapan model *multimodal* dengan model *DeiT-Small* dan *IndoBERT*.
2. Menyajikan evaluasi mengenai ketangguhan arsitektur *multimodal* berbasis *Feature-Level Fusion* dalam menghadapi kondisi ketika salah satu modalitas tidak tersedia, sehingga dapat memperkaya literatur terkait sistem deteksi konten ilegal.
3. Menjadi referensi bagi penelitian selanjutnya yang membahas klasifikasi iklan judi *online*, terutama ketika dihadapkan pada skenario gambar dengan teks yang disamarkan atau rusak.
4. Memberikan manfaat bagi pengguna *website* dalam melakukan pengawasan dan *filtering* konten khususnya dalam mengidentifikasi keberadaan konten judi *online*.

1.6 Kerangka Berpikir

Kerangka berpikir penelitian yang menggambarkan hubungan antara fakta, teori, masalah, solusi, metode, dan hasil yang diharapkan ditunjukkan pada Gambar 1.1.



Gambar 1.1 Kerangka berpikir penelitian

Berdasarkan Gambar 1.1, penelitian ini berangkat dari fakta bahwa penyebaran *website* judi *online* di Indonesia masih menjadi permasalahan serius. Hal ini diperkuat dengan data Kementerian Komunikasi dan Digital Republik Indonesia yang mencatat bahwa jutaan konten terkait judi *online* telah diblokir

hingga tahun 2025 [1]. Meskipun upaya pemblokiran terus dilakukan, iklan judi *online* masih dapat muncul kembali dalam berbagai bentuk promosi digital. Konten judi *online* memiliki karakteristik visual dan tekstual yang khas, seperti tampilan promosi yang mencolok, penggunaan istilah persuasif, serta penyamaran kata melalui bahasa slang atau tidak baku [4],[11]. Keberadaan konten tersebut berpotensi memberikan dampak negatif bagi masyarakat, terutama anak-anak dan remaja, baik dari sisi sosial, psikologis, finansial, maupun akademik [3],[6]. Oleh karena itu, diperlukan sistem deteksi otomatis yang mampu membantu proses penyaringan konten judi *online* pada *website*.

Dalam upaya membangun sistem deteksi yang efektif, pendekatan *multimodal deep learning* menjadi salah satu landasan yang relevan karena mampu memproses lebih dari satu jenis data secara bersamaan, seperti gambar dan teks [5]. Pendekatan ini telah digunakan dalam penelitian identifikasi *website* judi dengan memanfaatkan informasi visual dari tangkapan layar *website* dan informasi tekstual hasil *OCR* untuk meningkatkan performa klasifikasi [7]. Pada penelitian ini, model *Data-efficient Image Transformer* atau *DeiT-Small* digunakan untuk mengekstraksi fitur visual dari gambar karena arsitektur *DeiT-Small* dirancang sebagai model *vision transformer* yang efisien dalam mempelajari representasi visual [12]. Sementara itu, *IndoBERT* digunakan untuk mengekstraksi fitur tekstual karena merupakan model bahasa berbasis *BERT* yang dikembangkan untuk pemrosesan bahasa Indonesia [13].

Namun, berdasarkan kajian penelitian terdahulu, beberapa pendekatan *multimodal* masih menggunakan metode penggabungan pada tahap keputusan akhir atau *decision-level fusion*. Pendekatan seperti ini dapat menjadi kurang tangguh ketika salah satu modalitas mengalami gangguan. Pada konteks penelitian ini, gangguan tersebut dapat terjadi ketika *OCR* gagal membaca teks pada gambar akibat ukuran huruf kecil, resolusi rendah, kontras buruk, atau penggunaan *font* tidak umum [4],[9]. Kondisi tersebut sejalan dengan permasalahan *missing modality* atau *corrupted modality*, yaitu kondisi ketika salah satu modalitas tidak tersedia, rusak, atau tidak dapat diproses secara optimal sehingga performa model *multimodal* dapat menurun [10].

Sebagai solusi atas permasalahan tersebut, penelitian ini mengusulkan pendekatan *feature-level fusion*. Pendekatan ini dilakukan dengan menggabungkan representasi fitur visual dari *DeiT-Small* dan representasi fitur tekstual dari *IndoBERT* sebelum proses klasifikasi akhir. Dengan demikian, model tidak hanya mengambil keputusan berdasarkan hasil akhir dari masing-masing modalitas secara terpisah, tetapi mempelajari hubungan antara fitur visual dan fitur tekstual secara bersamaan [5],[7]. Untuk membuktikan ketangguhan pendekatan tersebut, penelitian ini juga menggunakan skenario *stress test* dengan mensimulasikan kondisi ketika informasi teks hasil *OCR* tidak tersedia. Hasil akhir yang diharapkan adalah sistem deteksi iklan judi *online* yang lebih adaptif dan lebih tangguh dibandingkan pendekatan *decision-level fusion*, terutama ketika salah satu modalitas mengalami gangguan [9], [10].

1.7 Sistematika Penulisan

Sistematika penulisan ini disusun untuk memberikan gambaran menyeluruh mengenai struktur penyusunan laporan tugas akhir. Bagian ini menguraikan muatan masing-masing bab serta alur logika yang menghubungkan setiap tahapan penelitian. Adapun laporan tugas akhir ini disusun dengan sistematika sebagai berikut.

BAB 1 PENDAHULUAN

Bab ini menguraikan konteks permasalahan terkait penyebaran iklan judi *online* yang manipulatif, rumusan masalah yang difokuskan pada perbandingan arsitektur *Feature-Level Fusion* dan *Decision-Level Fusion*, tujuan penelitian, batasan eksperimen, manfaat penelitian, kerangka berpikir, serta sistematika penulisan.

BAB 2 TINJAUAN PUSTAKA

Bab ini memuat tinjauan terhadap penelitian-penelitian terdahulu dan mempertegas gap penelitian. Selain itu, dijabarkan pula landasan teori yang mencakup *Deep Learning*, *Computer Vision*, *Natural Language Processing (NLP)*, arsitektur

Transformer DeiT-Small dan *IndoBERT*, mekanisme *Feature Fusion (Dense Layer)*, serta metrik evaluasi yang digunakan.

BAB 3 METODOLOGI PENELITIAN

Bab ini memaparkan alur dan prosedur penelitian yang diadaptasi dari kerangka kerja *Cross-Industry Standard Process for Data Mining (CRISP-DM)*. Pembahasan mencakup tahapan *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling* hingga *Evaluation*.

BAB 4 HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil dari serangkaian eksperimen dan pelatihan model yang telah dilakukan. Bagian pembahasan akan menganalisis secara komparatif mengenai performa dan tingkat ketangguhan (*robustness*) arsitektur *Feature-Level Fusion* dibandingkan *Decision-Level Fusion*, khususnya ketika model dievaluasi menggunakan skenario *stress test*.

BAB 5 SIMPULAN DAN SARAN

Bab ini merangkum simpulan akhir dari seluruh hasil eksperimen untuk menjawab rumusan masalah secara langsung. Pada bab ini juga disertakan saran-saran konstruktif sebagai rujukan untuk pengembangan sistem *multimodal* di masa mendatang