

BAB I

PENDAHULUAN

1.1 Latar Belakang

Fenomena judi *online* di Indonesia menunjukkan eskalasi pesat dan menjadi perhatian publik serta pengawas ruang digital. Pemerintah melaporkan penanganan lebih dari 1,3 juta konten judi *online* sejak awal 2025, dengan estimasi perputaran dana yang mencapai Rp 3,29 triliun per hari atau sekitar Rp 1.200 triliun setahun, naik signifikan dari Rp 981 triliun pada tahun 2024 [1],[2]. Popularitas fenomena ini diperparah oleh strategi pemasaran tersembunyi di media sosial, seperti iklan terselubung dalam konten viral, *giveaway* palsu, hingga penyamaran tautan [3].

Tren ini menegaskan perlunya analisis kuantitatif untuk memetakan pola percakapan dan modus promosi. Pemodelan topik merupakan pendekatan yang relevan untuk mengungkap struktur tematik dalam data teks berskala besar. *Latent Dirichlet Allocation (LDA)* telah banyak digunakan di Indonesia, misalnya untuk analisis tren ekonomi kreatif [4], pemetaan isu politik di *Twitter* [5], dan penelitian pendidikan [6].

Selain LDA, pendekatan *Nonnegative Matrix Factorization (NMF)* menawarkan alternatif berbasis dekomposisi matriks yang dikenal lebih interpretatif dan efisien [7]. Studi perbandingan terbaru menunjukkan keunggulan NMF dalam mengekstrak topik dari teks pendek maupun korpus besar [8]. Penelitian lanjutan juga mengusulkan variasi seperti *Deep NMF* [9], *constrained NMF* [10], hingga *Neural NMF* yang memungkinkan pembentukan struktur hierarkis multilapis [11].

Berdasarkan kondisi tersebut, penelitian ini bertujuan untuk menerapkan dan membandingkan LDA dan NMF dalam memetakan isu judi *online* di Indonesia. Fokus utama adalah mengidentifikasi topik promosi dan diskursus publik, menganalisis tren temporal, serta memberikan rekomendasi empiris yang dapat mendukung strategi mitigasi konten digital dan kebijakan pemerintah dalam memberantas judi *online*.

1.2 Rumusan Masalah

Berdasarkan uraian pada Latar Belakang Masalah, maka dapat dirumuskan beberapa permasalahan yang akan dikaji dalam penelitian ini, yaitu:

1. Topik apa saja yang dapat diidentifikasi dari opini publik terkait isu judi online di Indonesia menggunakan metode LDA dan NMF?
2. Sejauh mana metode LDA dan NMF mampu mengelompokkan komentar masyarakat secara koheren berdasarkan isu-isu yang muncul di media sosial *YouTube*?

1.3 Batasan Masalah

Untuk memperjelas ruang lingkup penelitian dan menghindari meluasnya pembahasan, maka penelitian ini memiliki batasan masalah sebagai berikut:

1. Data yang digunakan hanya berupa komentar publik dari platform *YouTube* yang terkait dengan isu judi *online* di Indonesia.
2. Proses pengolahan data dibatasi pada teks komentar dalam bahasa Indonesia, komentar dalam bahasa asing atau berbentuk non-teks (emoji, simbol, gambar) tidak dianalisis.
3. Metode analisis yang digunakan untuk pemodelan topik dibatasi pada dua pendekatan, yaitu *Latent Dirichlet Allocation (LDA)* dan *Non-Negative Matrix Factorization (NMF)*. Metode lain seperti *BERTopic* atau *transformer-based models* tidak dibahas secara mendalam.
4. Periode pengumpulan data dibatasi pada komentar yang diposting selama kurun waktu tertentu dari 30 Agustus 2023 hingga 7 Desember 2025, sehingga tidak mencakup seluruh data historis sejak awal kemunculan isu judi *online*.
5. Penelitian tidak membahas aspek hukum, sosial, maupun dampak ekonomi secara mendalam, melainkan hanya berfokus pada pemetaan topik di media sosial.

1.4 Tujuan Penelitian

Tujuan yang akan dicapai oleh peneliti dalam penelitian ini adalah sebagai berikut:

1. Mengidentifikasi topik-topik utama yang muncul dari opini publik terkait isu judi *online* di Indonesia menggunakan metode LDA dan NMF.
2. Mengevaluasi kualitas hasil pemodelan topik dengan menggunakan metrik seperti *coherence score* untuk menilai koherensi dan ketepatan topik yang terbentuk.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Manfaat Teoritis

1. Memberikan kontribusi pada pengembangan ilmu pengetahuan di bidang *Natural Language Processing (NLP)*, khususnya pada penerapan *topic modeling* terhadap isu sosial di Indonesia.
2. Menjadi referensi bagi penelitian selanjutnya yang ingin mengkaji fenomena sosial menggunakan data media sosial dengan pendekatan *Machine Learning*.

2. Manfaat Praktis

1. Memberikan gambaran mengenai persepsi publik di media sosial terkait isu judi *online*, yang dapat digunakan sebagai bahan evaluasi dalam merumuskan kebijakan atau program pencegahan.
2. Membantu masyarakat dan lembaga terkait memahami pola opini publik, sehingga dapat meningkatkan kesadaran dan literasi digital dalam menghadapi isu judi *online*.
3. Menyediakan dataset terstruktur yang dapat dimanfaatkan kembali oleh peneliti atau praktisi dalam penelitian lanjutan.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun dalam lima bab utama yang berisi uraian secara garis besar mengenai isi dari masing-masing bab, sebagai berikut:

1. **BAB I PENDAHULUAN:** Bab ini berisi latar belakang masalah, rumusan masalah, tujuan penelitian, batasan masalah, dan manfaat penelitian. Selain itu, bab ini juga menjelaskan urgensi dari penelitian terkait isu judi *online* di Indonesia, serta bagaimana penelitian ini dapat memberikan kontribusi bagi studi analisis media sosial, khususnya dalam pemodelan topik menggunakan *machine learning*.
2. **BAB II TINJAUAN PUSTAKA:** Bab ini menguraikan landasan teori yang mendukung penelitian, termasuk konsep dasar *text mining*, analisis sentimen, serta teori mengenai pemodelan topik. Tinjauan pustaka juga mencakup penelitian-penelitian terdahulu yang relevan, serta penjelasan tentang model dan algoritma yang digunakan dalam penelitian ini seperti LDA dan NMF untuk pemodelan topik, dan konsep NLP terkait (*tokenisasi*, *stopwords*, *word embedding*).
3. **BAB III METODE PENELITIAN:** Bab ini menjelaskan metode penelitian yang digunakan, meliputi objek penelitian, teknik pengumpulan data dari media sosial, tahap *preprocessing data* (*cleaning*, *case folding*, *stopword removal*, *stemming*, *tokenisasi*), serta tahapan implementasi model pemodelan topik. Selain itu, dijelaskan pula teknik evaluasi model yang digunakan, seperti *coherence score* untuk topik.
4. **BAB IV HASIL DAN PEMBAHASAN:** Bab ini menyajikan hasil penelitian, termasuk hasil pemodelan topik terkait isu judi *online* di Indonesia, hasil analisis sentimen publik terhadap isu tersebut, serta evaluasi performa model. Hasil yang diperoleh dianalisis lebih lanjut untuk memahami tren, persepsi publik, dan pola diskusi yang berkembang. Visualisasi data dalam bentuk grafik dan tabel ditampilkan guna mendukung interpretasi hasil penelitian.

5. **BAB V PENUTUP:** Bab ini berisi kesimpulan yang diperoleh dari penelitian sebagai jawaban atas rumusan masalah yang diajukan. Selain itu, bab ini memberikan saran-saran yang dapat menjadi rekomendasi untuk penelitian lebih lanjut, baik dalam pengembangan metode analisis, pemanfaatan data dari platform media sosial lain, maupun peningkatan akurasi model yang digunakan.

