

# BAB I PENDAHULUAN

## 1.1. Latar Belakang

Portal berita digital seperti Detik.com, Kompas.com, dan Tribunnews telah menjadi sumber informasi utama bagi masyarakat. Berdasarkan laporan Reuters Institute 2024, 50% responden di Indonesia mengakses Detik.com dalam sepekan, menjadikannya media online paling banyak diakses, diikuti oleh Kompas.com (39%) dan Tribunnews (28%) [1]. Tingginya kepercayaan publik terhadap portal-portal ini membuatnya menjadi platform penting dalam menyajikan isu-isu terkini, termasuk fenomena judi online yang sering menjadi sorotan. Meski demikian, tingginya volume artikel yang terus bertambah secara dinamis dari berbagai portal sering kali menyulitkan pembaca untuk menemukan informasi yang spesifik. Dalam konteks ini, penerapan text mining yang mengintegrasikan data dari beberapa portal berita menjadi solusi untuk mengolah informasi, mengekstrak lokasi, memetakan wilayah terdampak, serta mengklasifikasikan risiko yang terkandung dalam berita terkait judi online. Pendekatan berbasis text mining ini tidak hanya menghasilkan informasi yang lebih terstruktur [2], tetapi juga mendukung analisis berbasis data untuk memberikan wawasan yang lebih bermakna dalam memahami fenomena dan mendukung pengambilan keputusan.

Fenomena judi online menjadi permasalahan yang serius di Indonesia, tidak hanya melanggar norma sosial dan agama, tetapi juga berdampak signifikan pada perekonomian negara dengan aliran uang keluar secara *anonim* [3]. Berdasarkan laporan tahunan Pusat Pelaporan dan Analisis Transaksi Keuangan (PPATK), pada tahun 2023 tercatat 168 juta transaksi terkait judi online dengan nilai total mencapai Rp327 triliun, meningkat drastis dibandingkan tahun 2017 yang hanya 250 ribu transaksi senilai Rp2 triliun. Selain itu, sekitar 3,29 juta orang di Indonesia aktif terlibat dalam judi online, yang sering kali menggunakan rekening hasil penyalahgunaan atau jual beli untuk menyamarkan aktivitas ilegal tersebut. Modus ini diperparah dengan pengalihan dana ke luar negeri, menciptakan tantangan besar bagi penegak hukum [4]. Dalam konteks ini, penerapan teknologi seperti text mining menjadi salah satu upaya solusi, memungkinkan analisis berita untuk mengklasifikasikan risiko, dan memetakan wilayah terdampak[5]. Langkah ini tidak

hanya mendukung penegakan hukum, tetapi juga memberikan wawasan strategis untuk kebijakan yang lebih terarah.

Pemetaan lokasi dalam konteks judi online memiliki peran dalam memahami wilayah terdampak. Pemetaan lokasi yang dilakukan melalui pendekatan text mining memungkinkan visualisasi wilayah tertentu yang menjadi pusat resiko fenomena judi online[6]. Dengan menggunakan *Named Entity Recognition (NER)* untuk mengekstraksi informasi lokasi dari teks dan mengintegrasikannya dengan teknik pemetaan, analisis data menjadi lebih akurat dan terfokus [7]. Dalam kasus judi online, pemetaan lokasi dapat memberikan salah satu wawasan bagi penegak hukum untuk menentukan prioritas tindakan di wilayah yang memiliki risiko. Selain itu, pendekatan ini juga membantu pemerintah dalam menyusun kebijakan pencegahan dan penindakan berbasis data yang lebih terarah.

Text mining adalah teknik analisis data yang digunakan untuk mengekstrak informasi bermakna dari teks tidak terstruktur, seperti berita digital atau media sosial[8]. Dalam skripsi ini, text mining diterapkan untuk mengolah berita-berita terkait judi online. Algoritma Support Vector Machine (SVM) dimanfaatkan untuk mengklasifikasikan tingkat risiko. Penggunaan text mining memungkinkan proses analisis menjadi lebih efisien dibanding metode manual, serta menghasilkan pemetaan risiko yang akurat dan dapat digunakan sebagai dasar pengambilan keputusan yang lebih tepat sasaran[7].

Studi sebelumnya menunjukkan bahwa penggunaan *NER* dan *SVM* telah terbukti dalam analisis data teks tidak terstruktur, khususnya untuk mengekstrak dan mengklasifikasikan informasi geografis [7]. Dalam jurnal tersebut, *NER* digunakan untuk mengidentifikasi entitas lokasi dari teks Twitter, seperti nama kota, provinsi, atau tempat tertentu, yang kemudian dipetakan secara geografis. Sementara itu, *SVM* diterapkan untuk mengklasifikasikan jenis bencana berdasarkan pola teks, menghasilkan akurasi tinggi hingga 95%. Kombinasi kedua metode ini memungkinkan data tidak terstruktur, seperti teks media sosial atau berita, untuk diolah menjadi informasi yang terstruktur dan bermakna. Temuan ini memberikan dasar untuk penerapan metode serupa dalam penelitian lain, seperti analisis berita digital terkait fenomena judi online, untuk memetakan lokasi dan mengklasifikasikan risiko aktivitas ilegal tersebut secara lebih efisien [7].

Metode *NER* dan *SVM* dipilih dalam skripsi ini karena keduanya menawarkan pendekatan yang sesuai untuk pemetaan untuk menangani masalah analisis teks. *NER* digunakan untuk mengekstrak entitas penting dari teks tidak terstruktur, seperti nama, lokasi, dan organisasi, yang sangat relevan dalam mengidentifikasi wilayah dan pihak yang terkait dengan fenomena sosial, termasuk judi online[9]. Hasil ekstraksi ini kemudian diklasifikasikan menggunakan *SVM*, sebuah algoritma memiliki performa konsisten dalam tugas klasifikasi teks. *SVM* bekerja dengan mencari hyperplane optimal yang memaksimalkan margin antar kelas, memungkinkan pengklasifikasian risiko ke dalam kategori rendah, sedang, atau tinggi dengan tingkat akurasi tinggi. Studi sebelumnya menunjukkan bahwa *SVM* mencapai akurasi rata-rata 90,72%, menjadikannya metode yang efisien, terutama dalam menangani dataset besar[10]. Kombinasi *NER* dan *SVM* memastikan hasil analisis yang terstruktur, akurat, dan dapat diterapkan untuk mendukung pemetaan wilayah terdampak dan pengambilan keputusan.

Maka dari itu, tujuan skripsi ini adalah untuk menganalisis fenomena judi online melalui pendekatan text mining dengan memanfaatkan metode *NER* dan *SVM* untuk menghasilkan informasi yang terstruktur dan bermakna. Skripsi ini bertujuan untuk mengekstraksi entitas penting seperti lokasi dari berita digital, serta mengklasifikasikan tingkat risiko aktivitas judi online menjadi kategori rendah, sedang, dan tinggi. Selain itu, skripsi ini juga berupaya memetakan wilayah terdampak guna memberikan wawasan yang lebih mendalam tentang distribusi resiko judi online di Indonesia. Dengan pendekatan ini, skripsi diharapkan mampu mendukung penegakan hukum dan penyusunan kebijakan berbasis data, serta memberikan kontribusi salah satu upaya dampak sosial dan ekonomi yang disebabkan oleh aktivitas ilegal tersebut. Rentang waktu data yang digunakan dalam skripsi ini mencakup 2017 hingga 2024, berdasarkan laporan tahunan PPATK yang mencatat peningkatan tajam dalam transaksi judi online dan melibatkan sekitar 3,29 juta orang di Indonesia. Periode ini dipilih karena mencerminkan perkembangan fenomena judi online yang semakin marak dan dampaknya yang semakin signifikan terhadap masyarakat serta perekonomian negara.

## **1.2. Rumusan Masalah**

Berdasarkan latar belakang yang telah diuraikan sebelumnya, maka dapat dirumuskan beberapa permasalahan pada skripsi ini seperti berikut:

1. Bagaimana mengklasifikasikan lokasi judi online di Indonesia dengan menggunakan data berita ke dalam kategori risiko rendah, sedang, dan tinggi dengan menggunakan metode Support Vector Machine (SVM) dan Named Entity Recognition (NER) untuk mengekstrak lokasi dari teks berita?
2. Bagaimana memetakan judi online di Indonesia berdasarkan lokasi yang diekstraksi dari berita digital menggunakan metode text mining dan NER, serta menyajikan hasilnya dalam bentuk visualisasi yang dapat digunakan untuk mendukung pengambilan keputusan?

## **1.3. Batasan Masalah**

Batasan masalah yang digunakan dalam skripsi ini bertujuan untuk menghindari penyimpangan dalam pembahasan, sebagai berikut:

1. Data yang digunakan dari portal berita digital seperti Detik.com, Kompas.com, dan Tribunnews selama periode 2017 hingga 2024, yang berisi informasi fenomena judi online di Indonesia. Skripsi ini tidak mencakup data media sosial atau sumber lain di luar portal berita.
2. Skripsi berfokus pada pengklasifikasian tingkat risiko aktivitas judi online menjadi tiga kategori: rendah, sedang, dan tinggi. Skripsi tidak mencakup analisis faktor lain seperti profil demografi individu.
3. Pemetaan lokasi dilakukan berdasarkan klasifikasi risiko menggunakan data lokasi yang diekstraksi dari teks berita. Hasil ini terbatas pada representasi wilayah geografis di Indonesia yang diidentifikasi melalui berita digital dan tidak mencakup data lapangan.
4. Dalam pemetaan visual, apabila satu wilayah terdeteksi pada beberapa berita dengan kategori risiko yang berbeda, sistem akan menempatkannya di titik koordinat yang sama, sehingga dapat terjadi penumpukan atau hanya salah satu risiko yang ditampilkan.

#### **1.4. Tujuan Skripsi**

Berdasarkan latar belakang dan rumusan masalah yang telah disampaikan, maka tujuan dari skripsi ini adalah sebagai berikut:

1. Menganalisis tingkat risiko judi online di wilayah Indonesia dengan menggunakan metode Support Vector Machine (SVM).
2. Mengidentifikasi lokasi terdampak judi online di Indonesia melalui ekstraksi data menggunakan metode Named Entity Recognition (NER).

Memetakan hasil klasifikasi risiko judi online ke dalam kategori rendah, sedang, dan tinggi untuk mendukung pengambilan keputusan.

#### **1.5. Sistematika Penulisan**

Sistematika penulisan ini berisikan mengenai bagian-bagian yang akan diuraikan selama penulisan skripsi. Adapun hal ini dilakukan guna membantu penyusunan laporan supaya tidak menyimpang dari tujuan penulisan yang diharapkan. Tahapan dari penyusunan skripsi ini adalah sebagai berikut:

### **BAB I PENDAHULUAN**

Bab Pendahuluan berisikan mengenai gambaran umum skripsi yang akan dibuat. Bagian ini memiliki komposisi latar belakang, rumusan masalah, batasan masalah, tujuan, dan sistematika penulisan.

### **BAB II TINJAUAN PUSTAKA**

Bab Tinjauan Pustaka berisikan mengenai dasar teori dan penelitian terdahulu yang menjadi serupa dengan penelitian yang dilakukan. Dasar teori yang akan dijelaskan pada bagian ini berkaitan dengan

### **BAB III METODOLOGI PENELITIAN**

Bab ini menjelaskan mengenai metode penelitian yang akan digunakan, mencakup tahapan-tahapan seperti pengumpulan data, analisis kebutuhan, dan pengembangan model.

### **BAB IV HASIL DAN PEMBAHASAN**

Bab ini memaparkan penerapan metode yang telah dirancang, serta evaluasi dari uji coba skenario yang dilakukan. Hasil pengumpulan data akan dianalisis dan dibahas.

### **BAB V PENUTUP**

Bab ini berisi kesimpulan dari skripsi yang telah dilakukan, yang menjawab

semua pertanyaan yang diajukan dalam bab-bab sebelumnya.

### **DAFTAR PUSTAKA**

Pada bagian ini akan dijelaskan mengenai sumber-sumber literatur yang digunakan dalam penyusunan laporan skripsi ini, yang mencakup penjelasan dan pembahasan yang mendukung untuk memahami isi laporan.