

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi telah mempengaruhi berbagai aspek kehidupan manusia, termasuk dalam pengelolaan data yang kini beralih dari proses manual ke digital [1]. Peralihan ini meningkatkan efisiensi dan akurasi, serta memungkinkan kemajuan dalam berbagai bidang, seperti bisnis dan administrasi pemerintahan [2]. Namun, perkembangan teknologi ini juga menyebabkan ledakan data yang sangat besar dan bervariasi, yang memerlukan upaya pengelolaan dan analisis yang cermat untuk memperoleh informasi yang bermanfaat [3]. Salah satu tantangan utama dalam era digital ini adalah perlindungan data pribadi dan kejahatan siber, yang semakin meningkat seiring dengan banyaknya data yang dikumpulkan dan diproses [4].

Undang-Undang Informasi dan Transaksi Elektronik (UU ITE) adalah landasan hukum yang mengatur aktivitas di dunia digital, termasuk larangan terhadap konten yang melanggar kesusilaan, pencemaran nama baik, perjudian *online*, dan penyebaran berita bohong [5]. UU ITE juga memberikan dasar hukum untuk melawan kejahatan yang terjadi di dunia maya, yang semakin relevan mengingat tingginya tingkat pelanggaran yang terjadi [6].

Penegakan hukum terhadap pelanggaran UU ITE menghadapi tantangan besar, terutama terkait dengan krisis keadilan dan transparansi di sistem hukum Indonesia. Kelemahan dalam pengawasan dan penegakan hukum menyebabkan ketidakpercayaan masyarakat terhadap sistem hukum. Oleh karena itu, teknologi informasi dan kecerdasan buatan memiliki peran penting dalam meningkatkan transparansi dan akuntabilitas dalam penegakan hukum, terutama dalam kasus pelanggaran UU ITE. Implementasi teknologi dapat membantu memantau dan mengidentifikasi pola-pola tindak pidana yang berkaitan dengan UU ITE dengan lebih efektif [7].

Berdasarkan data survei yang dilakukan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) pada tahun 2023, jumlah pengguna internet di Indonesia mencapai 78,19%, atau sekitar 215.626.156 orang dari total populasi 275.773.901

jiwa. Angka ini terus meningkat setiap tahunnya [8]. Dalam siaran pers NO. 152/HM/KOMINFO/07/2023 yang diterbitkan oleh Biro Humas Kementerian Kominfo pada tahun 2023, tercatat 1.859 laporan terkait pemanfaatan rekening perbankan untuk aktivitas perjudian *online*. Angka tersebut merupakan bagian dari total 1.914 laporan yang diterima Kementerian Kominfo hingga tahun 2023. Hal ini menunjukkan tingginya tingkat kejahatan yang melanggar UU ITE. Kasus-kasus pelanggaran UU ITE yang dianalisis dalam penelitian ini mencakup 324 kronologi kejadian yang diambil dari situs web putusan Mahkamah Agung.

Klasifikasi adalah suatu metode dalam *machine learning* yang mengelompokkan data ke dalam kategori atau kelas tertentu berdasarkan fitur-fitur yang dimiliki [9]. *Text Mining* merupakan proses ekstraksi data dari teks, di mana sumber data umumnya berasal dari dokumen. Tujuan dari proses ini adalah untuk menemukan kata-kata yang dapat menggambarkan isi dokumen, sehingga memungkinkan dilakukan analisis hubungan antar dokumen tersebut [10]. Salah satu algoritma yang sering digunakan untuk klasifikasi dalam *text mining* adalah algoritma K-NN. *K-Nearest Neighbors* (K-NN) adalah algoritma pembelajaran mesin yang mengidentifikasi kelas atau kategori dari suatu data berdasarkan fitur-fitur yang ada dan mayoritas kelas dari tetangga terdekat data tersebut [11]. Untuk mengukur kedekatan antar data, metrik *Euclidean* sering digunakan. Metrik mengukur kedekatan antar data dengan menghitung jarak antara dua titik menggunakan rumus akar kuadrat dari jumlah kuadrat perbedaan nilai fitur yang bersesuaian. Metrik ini populer karena kesederhanaannya dan kemampuannya menangani data berdimensi tinggi [12]. K-NN memiliki keunggulan sebagai algoritma *non-parametric* yang tidak membuat asumsi tentang distribusi data. Keunggulannya mencakup toleransi terhadap gangguan dan efisiensi dalam menangani data besar. K-NN sering digunakan dalam klasifikasi karena kemudahan implementasi dan fleksibilitasnya [13], [14].

Penelitian berjudul "Penerapan *Text Mining* untuk Klasifikasi Jenis Pasal UU ITE Menggunakan Algoritma *Naive Bayes*" menunjukkan bahwa algoritma *Naive Bayes* efektif dalam mengklasifikasikan jenis pelanggaran UU ITE berdasarkan teks kronologi kasus, dengan akurasi mencapai 80%. Penelitian ini membuktikan

bahwa algoritma *Naive Bayes* dapat digunakan sebagai alat bantu untuk memberikan prediksi mengenai pasal UU ITE yang kemungkinan besar dilanggar, sekaligus meningkatkan kesadaran hukum masyarakat terhadap potensi pelanggaran dalam aktivitas *online* [15].

Performa algoritma K-NN dapat ditingkatkan dengan menerapkan beberapa metode, seperti SMOTE (*Synthetic Minority Oversampling Technique*) dan Dimensionality Reduction - PCA (*Principal Component Analysis*). SMOTE digunakan untuk menangani ketidakseimbangan data dengan menghasilkan sampel sintetis untuk kelas minoritas, sehingga pola lebih mudah dikenali [16]. PCA digunakan untuk mereduksi dimensi data sambil mempertahankan informasi penting, mengurangi kompleksitas, meningkatkan efisiensi, dan meminimalkan risiko *overfitting* [17], [18].

Berdasarkan uraian latar belakang masalah di atas maka penulis tertarik untuk melakukan penelitian dan menulis tugas akhir yang dengan judul “**Penerapan Text Mining Untuk Klasifikasi Jenis Pasal UU ITE Menggunakan Algoritma K-Nearest Neighbor (K-NN)**”, sebagai syarat untuk lulus pada mata kuliah Tugas Akhir serta dengan harapan dapat memberikan kontribusi dalam mengatasi tantangan penegakan hukum di era digital, khususnya terkait pelanggaran UU ITE. Penerapan teknologi *text mining* dan algoritma K-NN diharapkan mampu mempermudah identifikasi jenis pasal UU ITE yang dilanggar berdasarkan analisis teks putusan Mahkamah Agung. Dengan demikian, hasil penelitian ini diharapkan dapat menjadi landasan bagi pihak penegak hukum dan lembaga terkait untuk lebih efektif dan efisien dalam menangani kasus-kasus pelanggaran UU ITE.

1.2 Rumusan Masalah

Berdasarkan uraian masalah di atas, dapat disimpulkan bahwa permasalahan yang dihadapi dalam penelitian ini adalah:

1. Implementasi model klasifikasi *Text Mining* dengan algoritma *K-Nearest Neighbors* (K-NN) dalam konteks klasifikasi pasal-pasal Undang-Undang Informasi dan Transaksi Elektronik (UU ITE).

2. Evaluasi tingkat akurasi model klasifikasi tersebut dengan membandingkannya dengan data aktual yang terkait dengan kronologi pelanggaran pasal-pasal UU ITE.

1.3 Pertanyaan Penelitian

Berdasarkan rumusan masalah di atas, pertanyaan yang diajukan oleh peneliti dalam penelitian ini adalah:

1. Bagaimana cara mengimplementasikan model klasifikasi *Text Mining* dengan menggunakan algoritma *K-Nearest Neighbors* (K-NN) dalam konteks klasifikasi pasal-pasal Undang-Undang Informasi dan Transaksi Elektronik (UU ITE)?
2. Bagaimana melakukan evaluasi terhadap tingkat akurasi dari model klasifikasi tersebut dengan membandingkannya dengan data aktual yang terkait dengan kronologi pelanggaran pasal-pasal UU ITE?

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan penelitian ini dapat dijabarkan sebagai berikut:

1. Mengimplementasikan model klasifikasi *Text Mining* menggunakan algoritma *K-Nearest Neighbors* (K-NN) untuk mengklasifikasikan pasal-pasal Undang-Undang Informasi dan Transaksi Elektronik (UU ITE).
2. Mengevaluasi performa model klasifikasi *Text Mining* dalam mengidentifikasi jenis pelanggaran pasal-pasal UU ITE.

1.5 Batasan dan Asumsi Penelitian

Berdasarkan rumusan masalah dan pertanyaan penelitian, batasan dan asumsi penelitian ini dapat dijelaskan sebagai berikut:

1. Data berasal dari putusan Mahkamah Agung yang berkaitan dengan pelanggaran pasal 27 dan 28 Undang-Undang ITE nomor 11 tahun 2008 yang telah diubah dengan UU No. 19 Tahun 2016.
2. Algoritma akan mengklasifikasikan pasal 27 ayat 1, pasal 27 ayat 2, pasal 27 ayat 3, pasal 27 ayat 4, pasal 28 ayat 1 dan pasal 28 ayat 2.

3. Data yang akan digunakan dalam penelitian terbatas pada 324 dokumen kasus pelanggaran UU ITE yang termasuk dalam pasal 27 dan 28.
4. Sistem ini akan mengklasifikasi pasal 27 dan 28 UU ITE berdasarkan kronologi kasus yang dimasukkan oleh pengguna.
5. Algoritma yang digunakan dalam aplikasi ini adalah *K-Nearest Neighbors* (K-NN).

1.6 Manfaat Penelitian

Berdasarkan rumusan masalah, batasan masalah, dan tujuan penelitian yang telah diuraikan, manfaat dari penelitian ini dapat dijelaskan sebagai berikut:

1. Manfaat Teoritis
 - a. **Bagi Penulis:** Memperdalam pemahaman tentang K-NN, teknik Text Mining, dan pengembangan model klasifikasi dalam konteks hukum untuk referensi penelitian lanjutan.
 - b. **Bagi Masyarakat:** Memudahkan klasifikasi pelanggaran UU ITE, meningkatkan pemahaman hukum, dan kesadaran akan pelanggaran hukum digital.
 - c. **Bagi Universitas Telkom Purwokerto:** Menjadi referensi pengembangan teknologi Text Mining dan inspirasi bagi mahasiswa serta peneliti di bidang hukum digital.
2. Manfaat Praktis
 - a. **Bagi Penulis:** Meningkatkan keterampilan dalam menerapkan K-NN untuk klasifikasi teks dan memahami teknik seperti SMOTE, PCA, dan optimasi hyperparameter.
 - b. **Bagi Masyarakat:** Mempermudah klasifikasi kasus pelanggaran UU ITE, membantu memahami implikasi hukum aktivitas digital, dan memberikan wawasan tentang ketentuan hukum yang berlaku.
 - c. **Bagi Universitas Telkom Purwokerto:** Menjadi dasar pengembangan prototipe aplikasi hukum digital, mendorong kontribusi universitas dalam teknologi hukum, dan mendukung riset teknologi efisien di sektor hukum.