

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sistem peradilan di Indonesia menghasilkan jutaan dokumen hukum setiap tahunnya, terutama dari lembaga peradilan agama yang menangani perkara-perkara perdata seperti perceraian, hak asuh anak, warisan, hingga pembagian harta bersama. Jenis perkara tersebut secara konsisten mendominasi statistik perkara yang ditangani oleh peradilan agama di seluruh Indonesia. Berdasarkan data yang dipublikasikan secara daring melalui situs resmi Mahkamah Agung Republik Indonesia, hingga pertengahan tahun 2025 telah tercatat lebih dari 6,8 juta dokumen putusan perdata agama yang tersedia secara publik di laman putusan3.mahkamahagung.go.id, dengan lebih dari 5,4 juta di antaranya merupakan perkara perceraian [1]. Jumlah tersebut menunjukkan tingginya volume dokumen putusan perdata agama yang tersedia secara publik serta pentingnya pengelolaan dan dokumentasi hukum secara efektif dan transparan.



Gambar 1.1 Statistik Klasifikasi Dokumen Perdata Agama

Sumber: Direktori Putusan Mahkamah Agung [1]

Ketersediaan akses terbuka terhadap dokumen putusan tersebut merupakan bentuk penerapan prinsip keterbukaan informasi publik sebagaimana diatur dalam Undang-Undang Republik Indonesia Nomor 14 Tahun 2008 tentang Keterbukaan Informasi Publik [2]. Akses terbuka ini memungkinkan dokumen putusan dimanfaatkan oleh berbagai pihak, seperti mahasiswa, akademisi, praktisi hukum, dan masyarakat umum. Dokumen putusan perkara perdata agama juga merupakan salah satu objek yang banyak dimanfaatkan dalam penelitian hukum, khususnya pada bidang hukum keluarga Islam dan ekonomi syariah. Tingginya pemanfaatan dokumen tersebut sebagai objek kajian ilmiah bahkan mendorong lahirnya literatur khusus mengenai metode penelitian putusan pengadilan agama sebagai pedoman dalam menelaah dokumen sengketa secara sistematis [3]. Namun demikian, dokumen putusan pengadilan memiliki struktur yang panjang dan sistematis yang umumnya terdiri atas beberapa bagian utama, seperti identitas perkara, duduk perkara, pertimbangan hukum, hingga amar putusan. Dalam penyusunannya, uraian fakta, dalil para pihak, serta pertimbangan hukum disajikan secara rinci sehingga pada beberapa bagian terdapat pengulangan informasi [4]. Karakteristik tersebut menyebabkan proses penelusuran dan identifikasi informasi penting menjadi kurang efisien apabila dilakukan secara manual, terutama pada dokumen yang memiliki jumlah halaman yang banyak. Selain itu, variasi penggunaan kalimat dan istilah dalam dokumen hukum juga menjadi tantangan dalam proses identifikasi informasi penting secara cepat dan efisien [5].

Salah satu pendekatan yang dapat menjawab permasalahan tersebut adalah *automatic text summarization*, yaitu teknologi untuk menghasilkan ringkasan secara otomatis dari dokumen teks [6]. Pada dokumen hukum, *automatic text summarization* dapat membantu pembaca memperoleh informasi penting secara lebih cepat tanpa harus membaca keseluruhan isi dokumen yang umumnya panjang dan kompleks. Penerapan peringkasan otomatis juga dapat meningkatkan efisiensi dalam proses penelaahan dokumen hukum serta memudahkan berbagai pihak dalam memahami informasi penting yang terkandung di dalamnya [7]. Dalam bidang ini terdapat dua pendekatan utama, yaitu *extractive* dan *abstractive summarization*. Untuk konteks dokumen hukum, pendekatan *extractive* lebih sesuai digunakan karena ia mempertahankan kalimat asli dari dokumen sumber, sehingga dapat

menjaga keakuratan makna dan menghindari risiko penyimpangan yang dapat berdampak serius dalam konteks hukum [8]. Hal ini penting karena setiap kata dalam dokumen hukum mengandung makna yang dapat memengaruhi interpretasi dan putusan hukum itu sendiri.

Seiring perkembangan teknologi pemrosesan bahasa alami atau *Natural Language Processing* (NLP), telah banyak dikembangkan model-model berbasis *deep learning* untuk tugas *summarization*. Salah satu yang banyak digunakan dan diakui efektivitasnya adalah *BERTSum*, yaitu modifikasi dari arsitektur *BERT* (*Bidirectional Encoder Representations from Transformers*) yang dirancang khusus untuk peringkasan teks secara ekstraktif. Model ini mampu merepresentasikan setiap kalimat dalam dokumen sebagai vektor dan memilih kalimat-kalimat yang paling relevan untuk dimasukkan dalam ringkasan akhir [9]. Dalam penelitian Liu dan Lapata (2019) menunjukkan bahwa *BERTSum* berhasil mencapai skor ROUGE-1 sebesar 43,25, ROUGE-2 sebesar 20,24, dan ROUGE-L sebesar 39,63 pada dataset CNN/Daily Mail, yang lebih tinggi dibandingkan model *baseline* lainnya. Di Indonesia, telah dikembangkan pula model pralatih BERT khusus untuk Bahasa Indonesia, yaitu *IndoBERT*, yang dilatih menggunakan korpus Bahasa Indonesia dalam skala besar dan menunjukkan performa lebih baik dibanding *Multilingual BERT* (M-BERT) pada berbagai tugas NLP berbahasa Indonesia [10]. Dengan keunggulan tersebut, *IndoBERT* dapat dikombinasikan dengan arsitektur *BERTSum* dalam suatu pendekatan *hybrid*, di mana *IndoBERT* berperan sebagai representasi bahasa Indonesia, sedangkan *BERTSum* berfungsi sebagai mekanisme seleksi kalimat untuk menghasilkan ringkasan ekstraktif pada dokumen hukum berbahasa Indonesia.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk membangun sistem peringkasan otomatis dokumen putusan perkara perdata agama menggunakan pendekatan *extractive summarization* dengan kombinasi *IndoBERT* dan *BERTSum*. Untuk mengukur performa hasil ringkasan, digunakan metrik evaluasi *ROUGE* (*Recall-Oriented Understudy for Gisting Evaluation*), yang secara luas digunakan dalam penelitian *summarization* untuk mengevaluasi kualitas ringkasan berdasarkan kesamaan n-gram dengan ringkasan referensi [11]. Melalui penelitian ini, diharapkan sistem yang dibangun dapat memberikan ringkasan yang

relevan, informatif, dan efisien, sekaligus menjadi kontribusi nyata terhadap pengembangan teknologi hukum digital dan peningkatan akses masyarakat terhadap informasi hukum.

1.2 Rumusan Masalah

1. Bagaimana merancang sistem *automatic text summarization* berbasis kombinasi IndoBERT dan BERTSum untuk dokumen putusan hukum perdata agama berbahasa Indonesia?
2. Seberapa baik kombinasi IndoBERT dan BERTSum dalam menghasilkan ringkasan dokumen putusan perdata agama berdasarkan evaluasi ROUGE menggunakan *recall*, *precision*, dan F1-score?

1.3 Batasan Masalah

1. Penelitian ini difokuskan pada dokumen putusan perkara perdata agama yang diperoleh dari situs resmi Mahkamah Agung Republik Indonesia, dengan batasan dokumen putusan yang dipublikasikan pada tahun 2025.
2. Dari seluruh kategori perkara perdata agama yang tersedia, penelitian ini menggunakan beberapa jenis perkara, yaitu: perceraian (cerai gugat dan cerai talak), pengesahan nikah (isbat nikah), dispensasi nikah, kewarisan/harta waris, hibah dan wasiat, ekonomi syariah, izin poligami, pembatalan nikah, serta hak-hak setelah perceraian. Pemilihan jenis perkara tersebut dilakukan secara *purposive* untuk merepresentasikan variasi jenis perkara perdata agama, meskipun jumlah dokumen pada setiap jenis tidak seimbang.
3. Total dokumen yang digunakan dalam penelitian ini dibatasi sebanyak 30 dokumen putusan.
4. Metode peringkasan yang digunakan terbatas pada pendekatan *extractive summarization*, tanpa menggunakan pendekatan *abstractive*.
5. Model yang diterapkan adalah BERTSum dengan IndoBERT sebagai *encoder*, yang telah dilatih pada korpus bahasa Indonesia.
6. Evaluasi sistem ringkasan dilakukan menggunakan metrik ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*), yang meliputi ROUGE-1, ROUGE-2, dan ROUGE-L, dengan fokus pada F1-score sebagai ukuran utama performa ringkasan.

1.4 Tujuan

1. Merancang sistem *automatic text summarization* berbasis kombinasi IndoBERT dan BERTSum untuk dokumen putusan hukum perdata agama berbahasa Indonesia.
2. Mengevaluasi performa kombinasi IndoBERT dan BERTSum dalam menghasilkan ringkasan dokumen putusan perdata agama berdasarkan evaluasi ROUGE menggunakan *recall*, *precision*, dan F1-score.

1.5 Manfaat

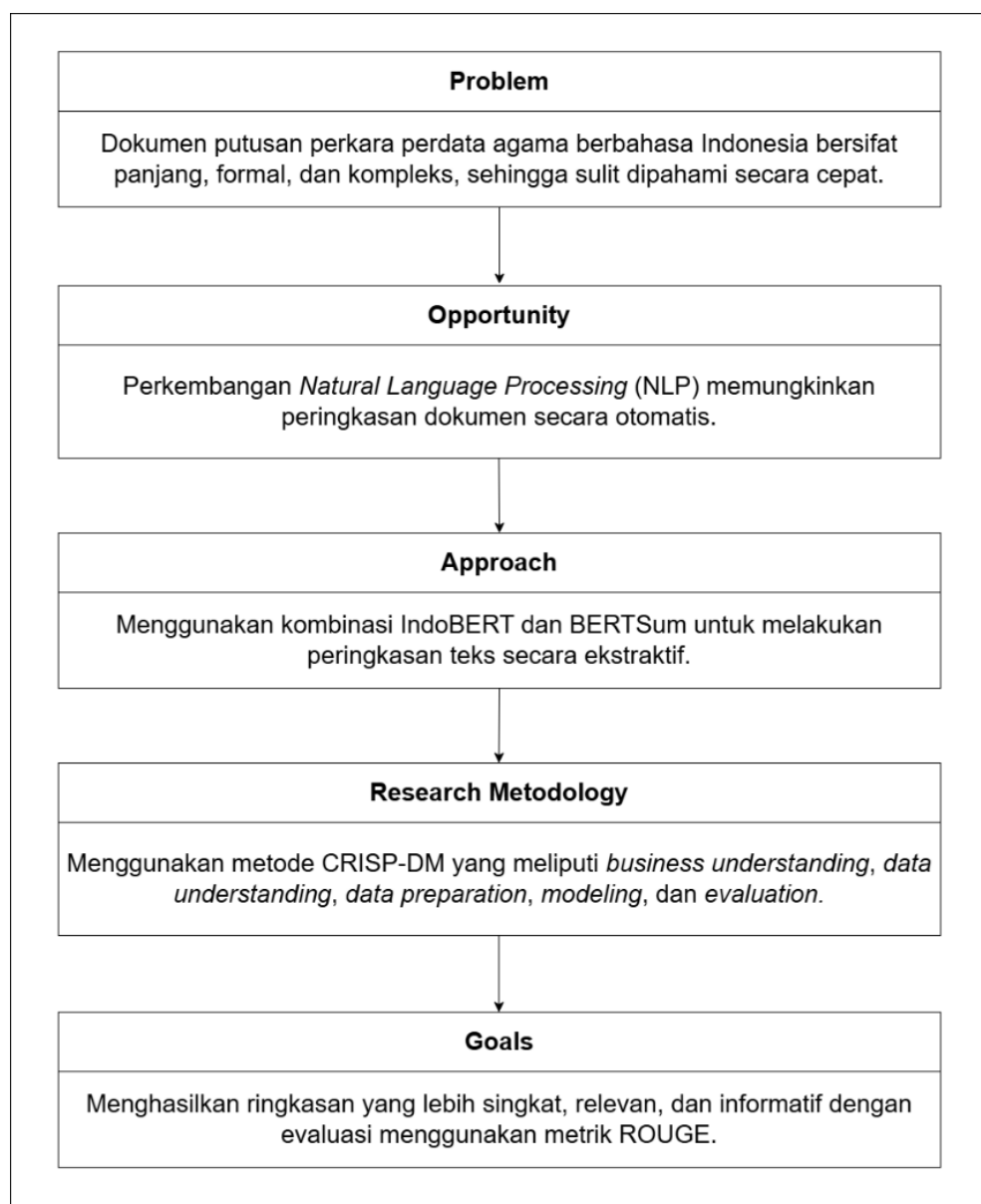
1. Menyediakan solusi teknologi berbasis *Natural Language Processing* (NLP) untuk meringkas dokumen hukum secara otomatis, sehingga dapat mempercepat akses terhadap informasi inti dari dokumen putusan yang panjang dan kompleks.
2. Meningkatkan efisiensi kerja profesional di bidang hukum dan akademik dengan menyajikan ringkasan yang singkat namun tetap akurat secara konteks hukum.
3. Mengembangkan dan mengevaluasi model *extractive summarization* berbasis IndoBERT, sebagai kontribusi dalam pengembangan teknologi NLP berbahasa Indonesia.

1.6 Kerangka Berpikir

Penelitian ini diawali dengan adanya permasalahan berupa sulitnya memahami dokumen putusan hukum perdata agama yang memiliki struktur panjang dan kompleks. Kondisi tersebut menyebabkan pembaca membutuhkan waktu yang relatif lama untuk menemukan informasi inti dari dokumen tersebut. Seiring dengan perkembangan teknologi, khususnya dalam bidang *Natural Language Processing* (NLP), muncul peluang untuk mengatasi permasalahan tersebut melalui teknik *automatic text summarization*, yang mampu menghasilkan ringkasan secara otomatis dari dokumen teks.

Dalam penelitian ini, digunakan pendekatan berbasis kombinasi IndoBERT dan BERTSum sebagai solusi untuk melakukan peringkasan teks secara ekstraktif. IndoBERT dimanfaatkan untuk memahami konteks bahasa Indonesia secara mendalam, sedangkan BERTSum digunakan untuk menentukan kalimat-kalimat

yang memiliki tingkat kepentingan tinggi dalam dokumen. Proses penelitian dilakukan menggunakan metode CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang meliputi tahapan *business understanding*, *data understanding*, *data preparation*, *modeling*, dan *evaluation*. Hasil dari penelitian ini diharapkan mampu menghasilkan ringkasan dokumen putusan hukum yang lebih singkat, relevan, dan informatif, serta memiliki performa yang baik berdasarkan evaluasi menggunakan metrik ROUGE.



Gambar 1.2 Kerangka Berpikir