**PROPOSAL SKRIPSI**

**SISTEM PENCARIAN DOKUMEN PUTUSAN PENGADILAN DENGAN RERANKER BERBASIS <<<*???>>>***



**Oleh:**

**M. SHOKHIBUL ANWAR**

**200411100130**

**Dosen Pembimbing 1 : Firdaus Solihin, S.Kom., M.Kom.**

**Dosen Pembimbing 2 :**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS TRUNOJOYO MADURA**

**2023**

# ABSTRAK

Jumlah dokumen putusan pengadilan yang dipublikasikan pada situs web resmi Mahkamah Agung Republik Indonesia terus bertambah dengan rata-rata 900.000 kasus per-tahun. Karenanya, penerapan sistem pencarian yang efektif sangat diperlukan untuk mempermudah dan mempercepat pencarian informasi relevan dari kasus terdahulu. Kendati demikian metode pencarian yang digunakan situs resmi Mahkamah Agung masih menggunakan metode pencarian tradisional *boolean search* yang menuntut pengguna untuk memiliki keahlian teknis yang tinggi agar dapat menggunakan sistem dengan maksimal. Penelitian ini bermaksud untuk menerapkan metode pencarian berbasis model probabilistik (BM25) yang digabung dengan *reranker* berbasis *Bidirectional Encoder Representations from Transformers (BERT).* Teknik BM25 mempertimbangkan frekuensi *term* dan karakteristik dokumen sehingga dapat meningkatkan fleksibilitas query serta menangani kompleksitas dan panjang dokumen. Sedangkan komponen *reranker* berfungsi untuk meningkatkan relevansi dokumen hasil BM25 dengan mempertimbangkan pemahaman semantik menggunakan model *pre-trained* berbasis transformer. Aplikasi metode ini diharapkan dapat meningkatkan akurasi dan fleksibilitas pencarian dokumen putusan pengadilan sehingga dapat meningkatkan efektifitas dan efisiensi pekerjaan para praktisi hukum.

Kata kunci : Information Retrieval, Legal Case Retrieval, BM25, BERT

# DAFTAR ISI

[ABSTRAK i](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756936)

[DAFTAR ISI ii](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756937)

[DAFTAR GAMBAR iv](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756938)

[DAFTAR TABEL v](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756939)

[BAB 1 1](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756940)

[1.1 Latar Belakang 1](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756941)

[1.2 Permasalahan 2](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756942)

[1.3 Metode Usulan 2](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756943)

[1.4 Pertanyaan penelitian 2](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756944)

[1.5 Tujuan 2](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756945)

[1.6 Manfaat 2](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756946)

[1.7 Batasan masalah 3](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756947)

[1.8 Sistematika Penulisan 3](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756948)

[BAB 2 4](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756949)

[2.1 Leukimia Limfoblastik akut (ALL) 4](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756950)

[2.2 Klasifikasi Citra 5](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756951)

[2.3 Pengolahan Citra Digital (PCD) 6](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756952)

[2.4 Augmentasi Data 7](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756953)

[2.5 Convolutional Neural Network (CNN) 8](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756954)

[2.5.1 *Convolutional Layer* 9](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756955)

[*2.5.2 Pooling Layer* 10](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756956)

[*2.5.3 Rectified Linear Unit (ReLU)* 11](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756957)

[*2.5.4 Fully Connected Layer (FCL)* 12](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756958)

[*2.5.5 Squeeze and Excitation (SE) Layer* 12](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756959)

[2.5.6 *Softmax* 14](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756960)

[2.6 HSV (Hue Saturation Value) 15](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756961)

[2.7 Fungsi aktivasi 16](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756962)

[2.7.1 Sigmoid 16](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756963)

[2.7.2 Swish 17](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756964)

[2.7.3 H – Swish 18](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756965)

[2.8 MobileNet 19](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756966)

[2.9 MobileNet V3 21](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756967)

[2.10 Confusion Matrix 22](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756968)

[2.11 Penelitian Terkait 24](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756969)

[BAB 3 27](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756970)

[3.1 Arsitektur Sistem 27](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756971)

[3.2 Dataset 27](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756972)

[3.3 Prepocessing 28](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756973)

[3.4 HSV Coloring 29](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756974)

[3.5 Mobile – Net V3 30](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756975)

[3.7 Skenario Pengujian 32](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756976)

[3.8 Evaluasi 33](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756977)

[3.9 Jadwal Penelitian 33](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756978)

[DAFTAR PUSTAKA 34](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756979)

# DAFTAR GAMBAR

[Gambar 2.1 Pembagian Class ALL 4](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262750)

[Gambar 2.2 Bentuk Leukimia Limfoblastik 5](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262751)

[Gambar 2.3 Tahapan CNN dalam mengklasifikasi 8](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262752)

[Gambar 2.4 Konvolusi antara matriks citra dan filter untuk baris 1 kolom 1 9](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262753)

[Gambar 2.5 Konvolusi antara matriks citra dan filter untuk baris 1 kolom 2 9](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262754)

[Gambar 2.6 Konvolusi antara matriks citra dan filter untuk baris 2 kolom 1 10](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262755)

[Gambar 2.7 Konvolusi antara matriks citra dan filter untuk baris 4 kolom 4 10](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262756)

[Gambar 2.8 A. Matriks dengan B. max-pooling dan C. average-pooling 11](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262757)

[Gambar 2.9 Operasi ReLU 11](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262758)

[Gambar 2.10 Ilustrasi Fully Connected Layer 12](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262759)

[Gambar 2.11 Gambaran dari HSV 15](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262760)

[Gambar 2.12 Sigmoid 16](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262761)

[Gambar 2.13 Tanh 17](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262762)

[Gambar 2.14 Arsitektur MobileNet 20](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262763)

[Gambar 2.15 Arsitektur MobileNet V3 21](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262764)

[Gambar 2.16 Contoh table Multiclass Confusion Matrix 23](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262765)

[Gambar 3.1 Arsitektur Program 27](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262766)

[Gambar 3.2 A)Citra Normal B)Citra HSV C)Hasil Masking 30](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262767)

# DAFTAR TABEL

[Tabel 2.1 Tabel Penelitian Terkait 25](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262835)

[Tabel 3.1 Tabel Jumlah data 28](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262836)

[Tabel 3.2 Tabel Persentasi Pembagian data 28](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262837)

[Tabel 3.3 Tabel Pembagian data Sebelum Augmentasi 28](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262838)

[Tabel 3.4 Tabel Pembagian data setelah Augmentasi 29](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262839)

[Tabel 3.5 Tabel Arsitektur Mobilenet-V3 31](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262840)

[Tabel 3.6 Tabel Skenario Pengujian 32](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262841)

[Tabel 3.7 Tabel jadwal penelitian 33](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147262842)

# 

**PENDAHULUAN**

## 1.1 Latar Belakang

Negara Kesatuan Republik Indonesia (NKRI) adalah sebuah negara hukum yang secara eksplisit ditegaskan melalui Undang-Undang Dasar Negara Republik Indonesia Pasal 1 ayat (3)[1]. Dalam penegakan hukum di Indonesia [2] menganut campuran antara alur proses buat putusan, data di ma, fungsi apa

Penjelasan putusan, karakteristik, kenapa perlu pencarian

Pencarian general, pencarian putusan, tantangan

Current pencarian di direktori, filtering, akurasi

Keunggulan boolean, kelemahan, solusi

Pretrained BERT, poin unggul, sekilas bukti keunggulan

Penelitian ini, hipotesis singkat, harapan

## 1.2 Permasalahan

Bagaimana menerapkan sebuah sistem pencarian informasi yang efektif dan efisien untuk mencari dokumen relevan dari basis data dokumen putusan pengadilan yang beragam serta, mengekstrak informasi spesifik dalam dokumen yang panjang dan memiliki struktur yang kompleks.

## 1.3 Metode Usulan

Implementasi model pencarian dengan algoritma pemeringkatan probabilistik berbasis *term* yaitu BM25 dengan komponen *reranker* berbasis *transfer learning* menggunakan model *Bidirectional Encoder Representations from Transformers (BERT)*  yang mampu menangkap makna semantik dari dokumen sehingga dapat meningkatkan akurasi pencarian.

## Pertanyaan penelitian

1. Seberapa akurat algoritma BM25 dalam menemukan dokumen relevan pada sistem pencarian dokumen putusan pengadilan?
2. Seberapa efektif penerapan BERT reranker dalam meningkatkan tingkat akurasi pencarian dokumen putusan pengadilan berbasis BM25?

## 1.5 Tujuan

Menerapkan metode pencarian BM25 dengan BERT reranker pada sistem pencarian dokumen putusan pengadilan Mahkamah Agung Republik Indonesia

## 1.6 Manfaat

Penelitian ini bertujuan memberikan kontribusi dalam bidang hukum khususnya terkait dokumen putusan pengadilan dengan menciptakan sistem pencarian yang lebih fleksibel dari segi query masukan dan memberikan hasil yang lebih akurat.

## 1.7 Batasan masalah

Penelitian ini berfokus pada :

1. Data yang digunakan berasal dari <https://github.com/ir-nlp-csui/indo-law> berisi 22630 dokumen putusan yang sudah dilakukan *parsing*, *cleaning* dan anotasi segmen oleh **(sitasi)**.
2. Data putusan yang hanya berjenis perkara pidana umum yang sudah berstatus berkekuatan hukum tetap *(inkracht)*.

## 1.8 Sistematika Penulisan

Struktur penulisan proposal disusun sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi 5 subbab yaitu: 1) Latar Belakang, 2) Rumusan Masalah, 3) Tujuan Penelitian, 4) Batasan Masalah, dan 5) Sistematika Penulisan. Menjelaskan konteks penelitian seperti pengenalan bidang penelitian, masalah yang ada dan kontribusi penelitian ini pada bidang tersebut.

BAB 2 LANDASAN TEORI

Dalam bab ini dijelaskan tentang teori-teori yang menjadi landasan bagi penelitian ini, yang mana berisikan: 1) Penjelasan *Information Retrieval*, 2.) *Legal Information Retrieval*, 3) *Best Matching 25* *(BM25)*, 4) *BERT*, dan 5) *BERT Reranker*.

BAB 3 METODOLOGI PENELITIAN

Bab ini menjelaskan metode penelitian yang akan diterapkan, mencakup analisis rancangan keseluruhan sistem, Pemrosesan dataset, alur penelitian, dan jadwal penelitian.

# 

**LANDASAN TEORI**

# DAFTAR PUSTAKA

[1] Indonesia, *Undang-Undang Dasar 1945, Pasal 1 Ayat (3)*. 1945.

[2] Indonesia, Pemerintah Pusat, *Undang-Undang No. 48 Tahun 2009*. 2009. Accessed: Feb. 24, 2024. [Online]. Available: http://peraturan.bpk.go.id/Details/38793/uu-no-48-tahun-2009