



PROPOSAL TUGAS AKHIR - ET234801

**SISTEM INFORMASI MANAJEMEN DASBOR
FILE DIGITAL DENGAN FILTERING TEXT DAN
PENCARIAN FILE BERBASIS KECERDASAN BUATAN**

MULKI KUSUMAH

NRP 05311940000043

Dosen Pembimbing

Dr. Ir. Henning Titi Ciptaningtyas, S.Kom., M.Kom.

NIP 19840708 201012 2 004

Program Studi S1 Teknologi Informasi

Departemen Teknologi Informasi

Fakultas Teknologi Elektro dan Informatika Cerdas

Institut Teknologi Sepuluh Nopember

LEMBAR PENGESAHAN

SISTEM INFORMASI MANAJEMEN DASBOR FILE DIGITAL DENGAN FILTERING TEXT DAN PENCARIAN FILE BERBASIS KECERDASAN BUATAN

PROPOSAL TUGAS AKHIR

Diajukan untuk memenuhi salah satu syarat
memperoleh gelar Sarjana Komputer pada
Program Studi S-1 Teknologi Informasi
Departemen Teknologi Informasi
Fakultas Teknologi Elektro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember

Oleh : Mulki Kusumah

NRP. 05311940000043

Disetujui oleh Tim Penguji Proposal Tugas Akhir :

1. Dr. Ir. Henning Titi Ciptaningtyas, S.Kom., (Pembimbing)
M.Kom.
NIP 19840708 201012 2 004

SURABAYA

Agustus, 2025

APPROVAL SHEET

PERNYATAAN ORISINALITAS

STATEMENT OF ORIGINALITY

ABSTRAK

SISTEM INFORMASI MANAJEMEN DASBOR FILE DIGITAL DENGAN FILTERING TEXT DAN PENCARIAN FILE BERBASIS KECERDASAN BUATAN

Nama Mahasiswa / NRP : **Mulki Kusumah / 05311940000043**
Departemen : **Teknologi Informasi FTEIC ITS**
Dosen Pembimbing : **Dr. Ir. Henning Titi Ciptaningtyas, S.Kom., M.Kom.**

Abstrak

Dalam era digital saat ini, kebutuhan akan sistem manajemen file yang efisien dan mudah diakses semakin meningkat, terutama dalam lingkungan kerja dan pendidikan. Penelitian ini bertujuan untuk merancang dan mengembangkan sebuah sistem informasi manajemen file digital berbasis web yang dilengkapi dengan fitur dasbor interaktif, unggah dan unduh file, serta pencarian file menggunakan filtering teks berbasis kecerdasan buatan (Artificial Intelligence/AI). Sistem ini memanfaatkan teknologi pemrosesan bahasa alami (Natural Language Processing/NLP) untuk meningkatkan akurasi dalam pencarian file berdasarkan isi atau kata kunci tertentu. Hasil pengujian menunjukkan bahwa sistem mampu mempercepat proses pencarian file dan meningkatkan efisiensi pengguna dalam mengelola dokumen digital. Dengan fitur filtering teks yang cerdas, pengguna dapat menemukan file dengan lebih tepat dan cepat, bahkan pada jumlah data yang besar. Sistem ini diharapkan dapat menjadi solusi efektif dalam manajemen dokumen digital yang aman, cepat, dan terorganisir.

Kata kunci: Sistem informasi, manajemen file digital, dasbor, pencarian file, filtering teks, kecerdasan buatan.

ABSTRACT

DIGITAL FILE DASHBOARD MANAGEMENT SYSTEM WITH AI BASED TEXT FILTERING AND FILE SEARCHING

Student Name / NRP : **Mulki Kusumah / 05311940000043**
Department : **Information Technology FTEIC-ITS**
Advisor : **Dr. Ir. Henning Titi Ciptaningtyas, S.Kom., M.Kom**

Abstract

In today's digital era, the need for efficient and accessible file management systems is increasing, especially in work and educational environments. This study aims to design and develop a web-based digital file management information system equipped with an interactive dashboard, file upload and download features, as well as AI-based text filtering and file search functionality. The system utilizes Natural Language Processing (NLP) to enhance the accuracy of file searches based on content or specific keywords. Testing results show that the system successfully accelerates the file search process and improves user efficiency in managing digital documents. With intelligent text filtering features, users can locate files more quickly and accurately, even within large datasets. This system is expected to provide an effective solution for secure, fast, and organized digital file management.

Keywords: *Information system, digital file management, dashboard, file search, text filtering, artificial intelligence*

DAFTAR ISI

LEMBAR PENGESAHAN	i
APPROVAL SHEET	ii
PERNYATAAN ORISINALITAS	iii
STATEMENT OF ORIGINALITY	iv
ABSTRAK	v
ABSTRACT	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	viii
DAFTAR TABEL	ix
DAFTAR SIMBOL	x
BAB 1 PENDAHULUAN	11
1.1 Latar Belakang	11
1.2 Rumusan Masalah	11
1.3 Batasan Masalah	11
1.4 Tujuan	11
1.5 Manfaat	11
BAB 2 TINJAUAN PUSTAKA	12
2.1 Hasil Penelitian Terdahulu	12
2.2 Dasar Teori	12
BAB 3 METODOLOGI	13
3.1 Metode yang digunakan	13
3.2 Bahan dan peralatan yang digunakan	13
3.3 Urutan pelaksanaan penelitian	13
BAB 4 HASIL DAN PEMBAHASAN	14
BAB 5 KESIMPULAN DAN SARAN	15
DAFTAR PUSTAKA	16
LAMPIRAN	17
BIODATA PENULIS	18

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR SIMBOL

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Di era digital saat ini, data dan informasi menjadi aset yang sangat berharga. Seiring dengan pesatnya perkembangan teknologi informasi, volume data yang harus dikelola semakin besar, yang menuntut sistem yang lebih efisien dalam mengelola, menyimpan, dan mengambil data tersebut. Salah satu bentuk informasi yang paling umum digunakan adalah file digital, yang sering kali disimpan dalam berbagai format dan ukuran. Namun, seiring bertambahnya jumlah file yang disimpan, pengelolaan dan pencarian file yang tepat menjadi semakin sulit dilakukan. Oleh karena itu, dibutuhkan sistem yang dapat memfasilitasi manajemen file secara efektif dan efisien.

Sistem informasi manajemen file digital menjadi solusi penting untuk memudahkan pengguna dalam mengelola, mencari, dan mengakses file secara cepat dan akurat. Namun, sebagian besar sistem yang ada masih memiliki keterbatasan dalam hal fungsionalitas, terutama dalam pencarian file berdasarkan isi atau kata kunci tertentu. Pada umumnya, pencarian file bergantung pada nama file atau metadata yang terbatas, yang menyebabkan pencarian menjadi tidak efektif ketika jumlah file yang dikelola sangat banyak.

Salah satu teknologi yang dapat meningkatkan kemampuan pencarian file adalah kecerdasan buatan (Artificial Intelligence/AI), khususnya pemrosesan bahasa alami (Natural Language Processing/NLP). Dengan menggunakan teknologi ini, sistem dapat mengidentifikasi dan memahami konten dari file secara lebih mendalam, memungkinkan pengguna untuk melakukan pencarian berdasarkan isi teks dalam file, bukan hanya berdasarkan nama atau metadata file. Hal ini tentu akan sangat membantu dalam meningkatkan akurasi dan efisiensi pencarian file, terutama dalam skala besar.

Sistem manajemen file berbasis web dengan dasbor interaktif yang dilengkapi dengan fitur upload, download, dan pencarian berbasis teks menggunakan AI diharapkan dapat memberikan solusi yang lebih efektif dalam manajemen dokumen digital. Pengguna dapat dengan mudah mengunggah dan mengunduh file, serta mencari file berdasarkan kata kunci tertentu dengan tingkat akurasi yang lebih tinggi. Sistem ini juga memungkinkan pengguna untuk melakukan filtering teks, sehingga pencarian file menjadi lebih spesifik dan relevan.

Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem informasi manajemen dasbor file digital dengan filtering teks dan pencarian file menggunakan kecerdasan buatan. Dalam penelitian ini, akan dibahas mengenai proses perancangan, teknologi yang digunakan, serta evaluasi dari sistem yang dibangun. Diharapkan, sistem ini dapat memberikan kontribusi yang signifikan dalam pengelolaan file digital yang lebih efektif dan efisien, serta memberikan kemudahan bagi pengguna dalam mengakses dan mencari dokumen digital mereka.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, permasalahan yang dihadapi dalam manajemen file digital adalah sebagai berikut:

1. Bagaimana merancang dan mengembangkan sistem informasi manajemen file digital berbasis web dengan fitur dasbor interaktif yang memungkinkan pengguna untuk mengunggah dan mengunduh file secara efisien?
2. Bagaimana mengimplementasikan fitur pencarian file yang efektif berdasarkan isi file (text-based search) menggunakan teknologi kecerdasan buatan (AI), khususnya pemrosesan bahasa alami (Natural Language Processing/NLP)?
3. Bagaimana sistem dapat menyediakan kemampuan filtering teks yang akurat untuk meningkatkan hasil pencarian dan mempermudah pengguna dalam menemukan file sesuai dengan kata kunci tertentu?
4. Apa saja tantangan dan solusi dalam mengelola file dengan volume besar dalam sistem manajemen file digital, serta bagaimana sistem ini dapat memastikan pengelolaan data yang aman dan terorganisir?

1.3 Batasan Masalah

Dalam penelitian ini, terdapat beberapa batasan masalah untuk memastikan fokus dan ruang lingkup yang jelas, antara lain:

1. Fokus pada pengelolaan file digital berbasis web, dengan fitur upload, download, dan pencarian file berbasis teks menggunakan kecerdasan buatan (AI).
2. Sistem ini hanya akan menangani file dalam format teks dan dokumen (misalnya .txt, .pdf, .docx) untuk fitur pencarian berbasis konten, dan tidak mencakup file multimedia (seperti gambar atau video).
3. Pencarian file berbasis teks menggunakan kecerdasan buatan hanya akan difokuskan pada penggunaan metode pemrosesan bahasa alami (Natural Language Processing/NLP) untuk analisis dan pencarian kata kunci dalam teks file.

1.4 Tujuan

Tujuan dari penelitian ini adalah untuk:

1. Merancang dan mengembangkan sebuah sistem informasi manajemen file digital berbasis web yang memungkinkan pengguna untuk mengunggah, mengunduh, dan mengelola file dengan mudah.
2. Mengimplementasikan fitur pencarian file berbasis teks menggunakan kecerdasan buatan, yang memungkinkan pengguna untuk menemukan file berdasarkan konten file dengan akurasi yang tinggi.
3. Mengembangkan fitur filtering teks yang dapat menyaring dan menampilkan file berdasarkan kata kunci yang relevan, meningkatkan efektivitas dan efisiensi pencarian file.
4. Mengidentifikasi tantangan dan memberikan solusi terkait pengelolaan file digital yang aman, efisien, dan terorganisir melalui sistem manajemen berbasis web.

1.5 Manfaat

Adapun manfaat dari penelitian ini yaitu :

Sistem yang dikembangkan akan memberikan kemudahan bagi pengguna dalam mengunggah, mengunduh, dan mengelola file digital secara terpusat, sehingga dapat mengurangi kompleksitas dalam pengelolaan file yang tersebar di berbagai tempat.

Dengan menggunakan teknologi pencarian berbasis teks yang didukung kecerdasan buatan (AI) dan pemrosesan bahasa alami (NLP), sistem ini dapat membantu pengguna untuk menemukan file berdasarkan isi teks dengan lebih cepat dan akurat, meskipun jumlah file yang dikelola sangat besar.

Fitur filtering teks yang akurat memungkinkan pencarian file yang lebih spesifik berdasarkan kata kunci atau frasa yang relevan, meningkatkan efisiensi pengguna dalam menemukan file yang sesuai dengan kebutuhan mereka.

Sistem ini akan memastikan pengelolaan file yang lebih terorganisir dan aman, memungkinkan pengguna untuk mengakses file dengan lebih efisien dan meminimalkan kesalahan dalam pengelolaan dokumen digital.

BAB 2 TINJAUAN PUSTAKA

2.1 Hasil Penelitian Terdahulu

Penelitian ini didasari pada beberapa hasil penelitian terdahulu sebagai pedoman dan referensi dalam proses penelitian ini. Informasi yang disampaikan dalam Tabel 2.1 berisi informasi mengenai penelitian sebelumnya yang berhubungan dengan penelitian ini

No	Tahun, Penulis	Pembahasan
1	Pembuatan Web Dashboard Interaktif untuk Analisis Data Penjualan Vending Machine	
	Tsani, Ajar Basyar, et al., 2025	Penelitian ini membahas pengembangan sebuah web dashboard interaktif yang dirancang untuk menganalisis data penjualan pada vending machine. Dashboard ini bertujuan untuk mempermudah pengelolaan dan visualisasi data dalam jumlah besar dengan menyediakan fitur-fitur interaktif seperti filter, pencarian, dan pagination. Melalui penggunaan teknologi open-source seperti HTML, CSS, JavaScript, Chart.js, dan Grid.js, dashboard ini dapat diakses secara efisien melalui GitHub Pages, sehingga memudahkan pengguna dalam memantau tren penjualan, produk terlaris, serta preferensi metode pembayaran. Hasil dari pengembangan ini menunjukkan bahwa sistem yang dibangun mampu memberikan informasi yang relevan untuk mendukung pengambilan keputusan strategis berbasis data. Namun, penelitian ini masih memiliki keterbatasan, terutama dalam hal belum terintegrasinya analitik prediktif yang dapat memberikan wawasan lebih mendalam, seperti proyeksi penjualan atau rekomendasi inventori. Oleh karena itu, pengembangan di masa depan disarankan untuk menambahkan algoritma prediksi dan menguji performa sistem pada data berskala besar. Secara keseluruhan, dashboard ini memberikan kontribusi nyata dalam optimalisasi pengelolaan vending machine dan memiliki potensi untuk diterapkan pada

		berbagai sektor bisnis lainnya.
2	<i>The application of NLP in information retrieval</i>	
	Wang Xurui., 2023	<p>Penelitian ini adalah tentang penerapan <i>Natural Language Processing (NLP)</i> dalam sistem temu kembali informasi (<i>information retrieval</i>). Fokus penelitian ini adalah bagaimana perkembangan NLP mampu meningkatkan relevansi dan kualitas hasil pencarian dibandingkan metode pencocokan kata kunci tradisional. Data yang digunakan berasal dari berbagai studi penerapan NLP di beberapa bidang, antara lain akademik, medis, pariwisata, dan e-commerce. Metode yang dibahas meliputi <i>document similarity-based algorithm</i> yang menggunakan ukuran kesamaan teks seperti <i>cosine similarity</i>, <i>Named Entity Recognition (NER)</i> untuk mengenali entitas penting dalam teks, serta perbandingan berbasis model representasi teks modern yang mengandalkan embedding vektor. Penelitian ini juga menguraikan implementasi NLP pada sistem pencarian literatur akademik, pembangunan <i>knowledge graph</i> dan sistem tanya-jawab medis, personalisasi rekomendasi destinasi wisata, analisis sentimen ulasan pariwisata, serta rekomendasi produk dan analisis ulasan dalam e-commerce. Berdasarkan hasil pembahasan, tren masa depan sistem <i>information retrieval</i> dengan NLP akan berfokus pada pemahaman semantik yang lebih dalam, pencarian yang dipersonalisasi, serta integrasi <i>multimodal retrieval</i> yang menggabungkan teks, gambar, dan suara sebagai sumber informasi (Wang, 2023).</p>
3	<i>Document Management System as an Integration Basis for Forming a Unified Data Source of Information Support of a Product Life Cycle</i>	
	Vilkov, Y. V., Chernysh, B. A., Kartamyshev, A. S., et al., 2019.	<p>Penelitian ini adalah tentang pengembangan <i>Document Management System (DMS)</i> sebagai dasar integrasi dalam membentuk sumber data terpadu untuk mendukung informasi sepanjang <i>product life cycle</i>. Sistem yang dikembangkan bertujuan untuk</p>

		<p>mengatasi fragmentasi data yang biasanya terjadi dalam siklus hidup produk, di mana informasi tersebar pada berbagai departemen dan aplikasi berbeda. Penelitian ini menggunakan pendekatan integrasi yang menjadikan DMS sebagai pusat pengelolaan, sehingga setiap data yang terkait dengan produk dapat tersimpan, diakses, dan dimanfaatkan secara konsisten di seluruh tahapan siklus hidup, mulai dari desain, produksi, hingga pemeliharaan.</p> <p>Metode yang dibahas menekankan pada penggunaan DMS sebagai platform utama integrasi data, dengan kemampuan menghubungkan berbagai sistem informasi lain yang sudah ada di perusahaan. Dengan demikian, DMS tidak hanya berfungsi sebagai penyimpanan dokumen, tetapi juga sebagai penghubung yang memastikan keselarasan informasi antar unit kerja. Penelitian ini menguraikan bagaimana sistem ini mampu meningkatkan efisiensi, mengurangi duplikasi data, serta memperbaiki kualitas informasi yang digunakan dalam pengambilan keputusan terkait produk.</p>
--	--	---

2.2 Dasar Teori

2.2.1 *Natural Language Processing*

Natural Language Processing (NLP) merupakan cabang dari kecerdasan buatan (Artificial Intelligence) yang berfokus pada interaksi antara komputer dan bahasa manusia. NLP memungkinkan sistem untuk memahami, memproses, dan menghasilkan teks atau ucapan dalam bahasa alami.

Dalam konteks pengelolaan file digital, NLP dapat digunakan untuk melakukan *text preprocessing* seperti tokenisasi, stemming, dan penghapusan *stopwords*, sehingga mempermudah proses analisis isi dokumen. Selain itu, NLP juga mendukung fitur pencarian cerdas (*intelligent search*) dengan cara memahami makna kata, bukan sekadar pencocokan

literal. Menurut Jurafsky dan Martin (2021), NLP mencakup serangkaian teknik yang memungkinkan komputer untuk memahami, menginterpretasikan, dan menghasilkan bahasa manusia.

Dalam pengembangan sistem informasi, NLP digunakan untuk:

1. **Text Preprocessing** – tahap awal yang meliputi *tokenization*, *stopword removal*, *stemming*, dan *lemmatization*. Tahap ini bertujuan untuk menyiapkan teks agar dapat dianalisis secara lebih efektif.
2. **Text Representation** – mengubah teks menjadi bentuk numerik, misalnya *Bag-of-Words*, *TF-IDF*, atau representasi vektor dengan *word embeddings* (Word2Vec, GloVe, BERT).
3. **Text Classification** – mengelompokkan teks ke dalam kategori tertentu, misalnya filtering konten, spam detection, atau sentiment analysis.

Dalam konteks penelitian ini, NLP berperan penting untuk melakukan *filtering* dan pencarian berbasis teks, sehingga sistem dapat menemukan file sesuai query pengguna dengan pemahaman yang lebih mendalam terhadap konten dokumen.

2.2.2 Information System Dashboard

Dashboard sistem informasi adalah antarmuka visual yang menyajikan informasi penting secara ringkas, interaktif, dan mudah dipahami untuk mendukung proses pengambilan keputusan. Dashboard biasanya menampilkan indikator, grafik, tabel, maupun *alerts* yang terhubung dengan sistem data di belakangnya.

Dalam pengelolaan file digital, dashboard berperan untuk memberikan kemudahan bagi pengguna dalam memantau, mengakses, dan mengelola file secara terpusat. Prinsip desain dashboard meliputi kejelasan, konsistensi, serta penyajian informasi yang relevan sesuai kebutuhan pengguna.

2.2.3 Text filtering

Text filtering adalah proses penyaringan teks untuk memilih atau menghapus konten tertentu berdasarkan aturan atau kriteria yang ditentukan. Filtering dapat dilakukan dengan metode berbasis aturan (*rule-based filtering*) atau berbasis pembelajaran mesin (*machine learning-based filtering*).

Beberapa teknik text filtering meliputi:

1. **Keyword Matching** – pencarian berdasarkan kata kunci tertentu.

2. ***Regular Expressions (Regex)*** – pencocokan pola teks yang lebih kompleks.
3. ***Machine Learning Filtering*** – menggunakan algoritma klasifikasi teks seperti Naïve Bayes, SVM, atau Neural Networks untuk memisahkan konten relevan dan tidak relevan.
4. ***Context-Aware Filtering*** – menggunakan NLP dan *semantic analysis* untuk mempertimbangkan makna konteks, bukan sekadar pencocokan literal.

Penerapan text filtering banyak ditemukan pada *spam filter email*, moderasi konten di media sosial, hingga sistem pencarian dokumen akademik. Dalam penelitian ini, *text filtering* digunakan untuk menyaring file sesuai kata kunci atau kategori yang dicari oleh pengguna.

2.2.4 *Information Retrieval*

Information Retrieval (IR) merupakan salah satu bidang dalam ilmu komputer yang berfokus pada pencarian informasi dari koleksi data dalam jumlah besar. Menurut Manning, Raghavan, dan Schütze (2008), IR adalah proses menemukan kembali informasi yang relevan dari sebuah koleksi dokumen berdasarkan permintaan (*query*) pengguna. Berbeda dengan sistem basis data yang biasanya menggunakan pencarian berbasis nilai yang pasti, IR bersifat tidak pasti karena sering kali hasil yang ditampilkan diurutkan berdasarkan tingkat relevansinya, bukan hanya pencocokan literal. Hal ini membuat IR sangat relevan dalam konteks pencarian dokumen teks atau file digital yang jumlahnya banyak.

Komponen utama IR meliputi:

- *Indexing* adalah membangun indeks kata atau metadata dokumen agar pencarian lebih cepat.
- *Query Processing* adalah mengolah input pengguna menjadi bentuk yang dapat dibandingkan dengan indeks.
- *Ranking* adalah mengurutkan hasil pencarian berdasarkan relevansi dengan *query* pengguna.

Dalam penerapannya, terdapat beberapa model IR yang sering digunakan. Model Boolean Retrieval menggunakan logika Boolean (AND, OR, NOT) untuk menemukan dokumen yang sesuai. Model Vector Space memperlakukan dokumen dan *query* sebagai vektor dalam ruang multi-dimensi, di mana kesesuaian dihitung dengan metode seperti *cosine similarity*. Sementara itu, model probabilistik mencoba menghitung probabilitas relevansi sebuah dokumen terhadap *query* tertentu. Ketiga model ini memberikan dasar bagi berbagai teknik pencarian modern, termasuk yang digunakan dalam sistem manajemen file berbasis teks.

Dalam penelitian ini, IR menjadi komponen utama yang mendasari pencarian file berbasis isi teks. Dengan menggunakan IR, sistem dapat menampilkan hasil pencarian file yang relevan berdasarkan *query* pengguna, bukan hanya mencocokkan nama file. Hal ini memberikan

efisiensi dan akurasi lebih tinggi dalam menemukan informasi yang dibutuhkan pengguna dari koleksi file digital yang besar.

2.2.5 *Storage Hierarchy*

Storage hierarchy adalah konsep dalam sistem komputer yang menjelaskan tingkatan media penyimpanan berdasarkan kecepatan akses, kapasitas, serta biaya per satuan penyimpanan. Menurut Tanenbaum (2015), hierarki penyimpanan dibangun atas dasar prinsip bahwa semakin cepat media penyimpanan, semakin kecil kapasitasnya dan semakin mahal biayanya. Oleh karena itu, sistem komputer menggunakan kombinasi dari berbagai tingkat penyimpanan untuk mencapai keseimbangan antara kinerja, kapasitas, dan biaya.

Lapisan tertinggi dalam hierarki adalah register, yang terdapat langsung di dalam prosesor. *Register* memiliki kecepatan akses tertinggi namun kapasitasnya sangat kecil. Di bawahnya terdapat *cache memory*, yang berfungsi untuk menyimpan data yang sering digunakan agar prosesor dapat mengaksesnya dengan cepat. Selanjutnya terdapat *main memory* (RAM), yang memiliki kapasitas lebih besar dibanding *cache* namun lebih lambat. Lapisan berikutnya adalah *secondary storage*, seperti *hard disk drive* (HDD) atau *solid-state drive* (SSD), yang digunakan untuk menyimpan data permanen. Lapisan terendah adalah *tertiary storage*, seperti *tape drive*, *optical disk*, atau *cloud storage*, yang digunakan untuk penyimpanan arsip jangka panjang.

Konsep hierarki penyimpanan ini sangat relevan dalam konteks sistem manajemen file digital. File yang sering diakses dapat disimpan pada media dengan kecepatan tinggi seperti SSD atau *cache*, sedangkan file yang jarang digunakan dapat dipindahkan ke penyimpanan dengan kapasitas besar tetapi lebih lambat seperti *cloud storage*. Dengan demikian, efisiensi sistem dapat ditingkatkan tanpa mengorbankan biaya. Selain itu, strategi penyimpanan ini mendukung skalabilitas sistem ketika jumlah *file* yang dikelola semakin besar.

Dalam penelitian ini, pemahaman tentang *storage hierarchy* digunakan untuk merancang sistem manajemen file yang optimal, khususnya dalam mengatur lokasi penyimpanan file serta strategi *indexing* untuk mendukung pencarian cepat. Dengan memanfaatkan prinsip hierarki penyimpanan, sistem yang dibangun diharapkan dapat menyeimbangkan performa pencarian dengan efisiensi penggunaan sumber daya penyimpanan.

2.2.6 *TF-IDF*

TF-IDF merupakan salah satu metode pembobotan kata yang banyak digunakan dalam text mining, machine learning, dan information retrieval. Menurut Salton dan McGill (1983), TF-IDF digunakan untuk mengevaluasi seberapa penting suatu kata terhadap sebuah dokumen dalam koleksi atau korpus. Inti dari metode ini adalah memberikan bobot lebih besar pada kata yang sering muncul dalam dokumen tertentu namun jarang muncul di dokumen lain.

Konsep TF-IDF terdiri dari dua komponen utama. Pertama, *Term Frequency* (TF) yang mengukur seberapa sering sebuah kata muncul dalam dokumen tertentu. TF biasanya dinormalisasi dengan membagi jumlah kemunculan kata dengan total kata dalam dokumen.

Kedua, *Inverse Document Frequency* (IDF) yang mengukur keunikan kata dengan membandingkan jumlah total dokumen dengan jumlah dokumen yang mengandung kata tersebut. Kata yang muncul di banyak dokumen, seperti kata hubung atau kata umum (“dan”, “atau”, “adalah”), akan memiliki nilai IDF rendah sehingga bobotnya kecil dalam pencarian.

Secara matematis, bobot TF-IDF untuk kata t dalam dokumen d dirumuskan sebagai berikut:

$$TF(t, d) = \frac{f_{t,d}}{\sum_k f_{k,d}}$$

$$IDF(t) = \log \frac{N}{df(t)}$$

$$TF-IDF(t, d) = TF(t, d) \times IDF(t)$$

di mana $f_{t,d}$ adalah jumlah kemunculan kata t dalam dokumen d , $df(t)$ adalah jumlah dokumen yang mengandung kata t , dan N adalah jumlah total dokumen dalam korpus.

Dalam sistem pencarian file, TF-IDF sangat efektif untuk mengukur relevansi antara query pengguna dengan isi file. Dengan memberikan bobot tinggi pada kata-kata penting, sistem dapat menampilkan hasil pencarian yang lebih akurat. Oleh karena itu, TF-IDF menjadi komponen penting dalam penelitian ini untuk mendukung pencarian file berbasis isi teks.

2.2.7 Content-Based Filtering

Content-based filtering adalah salah satu pendekatan dalam sistem rekomendasi yang berfokus pada analisis konten dari item yang tersedia. Menurut Ricci et al. (2011), content-based filtering bekerja dengan cara mengekstraksi fitur-fitur dari item (misalnya kata kunci, kategori, atau metadata) kemudian membandingkannya dengan kebutuhan atau preferensi pengguna. Pendekatan ini tidak membutuhkan data dari pengguna lain seperti pada *collaborative filtering*, sehingga cocok digunakan dalam sistem pencarian file yang lebih individual.

Tahapan utama content-based filtering meliputi ekstraksi fitur, perhitungan kesamaan, dan rekomendasi. Pada tahap ekstraksi fitur, sistem mengidentifikasi karakteristik unik dari dokumen atau file, seperti kata kunci dominan atau kategori isi. Tahap berikutnya adalah

menghitung kesamaan antar dokumen dengan menggunakan metode seperti cosine similarity atau Euclidean distance. Hasilnya kemudian digunakan untuk merekomendasikan dokumen yang paling mirip dengan query atau file yang sedang dicari.

Salah satu keunggulan content-based filtering adalah kemampuannya untuk menghasilkan rekomendasi yang bersifat personal dan konsisten, karena hanya bergantung pada konten file itu sendiri. Namun, pendekatan ini juga memiliki keterbatasan, misalnya *overspecialization*, di mana sistem hanya menampilkan file yang sangat mirip dan mengabaikan variasi lain yang mungkin juga relevan. Untuk mengatasi hal ini, beberapa penelitian menggabungkan content-based filtering dengan teknik lain seperti *collaborative filtering* untuk membentuk sistem rekomendasi hibrida.

Dalam penelitian ini, content-based filtering digunakan sebagai dasar dalam proses pencarian file yang relevan dengan query pengguna. Dengan menganalisis konten file melalui teknik NLP dan pembobotan TF-IDF, sistem dapat memberikan hasil pencarian yang lebih akurat dan sesuai dengan kebutuhan pengguna.

2.2.8 Database Management System

Database Management System (DBMS) adalah perangkat lunak yang digunakan untuk mendefinisikan, membuat, mengelola, dan mengakses basis data. Menurut Silberschatz, Korth, dan Sudarshan (2020), DBMS menyediakan antarmuka bagi pengguna untuk berinteraksi dengan data secara terstruktur, sekaligus menjaga integritas, keamanan, dan konsistensi data. DBMS memungkinkan pengelolaan data dalam jumlah besar secara efisien, baik untuk data terstruktur maupun semi-terstruktur.

Komponen utama dalam DBMS meliputi *Data Definition Language* (DDL) untuk mendefinisikan struktur basis data, *Data Manipulation Language* (DML) untuk mengakses dan memodifikasi data, serta query processor yang mengeksekusi perintah pengguna. Selain itu, DBMS juga memiliki storage manager yang bertanggung jawab terhadap penyimpanan fisik data, serta sistem indexing untuk mempercepat proses pencarian. Dalam konteks manajemen file, DBMS sangat penting karena mampu mengelola metadata file, seperti nama file, lokasi penyimpanan, ukuran, dan atribut tambahan yang digunakan dalam pencarian.

Salah satu fitur penting DBMS adalah indexing, yang memungkinkan pencarian data dilakukan lebih cepat dibandingkan pencarian linear. Dengan indeks, sistem dapat langsung mengakses lokasi data tanpa harus memeriksa seluruh isi tabel. Hal ini sangat bermanfaat ketika berhadapan dengan jumlah file yang besar, karena mempercepat proses pencarian berdasarkan kata kunci atau atribut file. DBMS modern juga mendukung pencarian full-text, yang memungkinkan pencarian dilakukan langsung pada isi dokumen, bukan hanya metadata.

Dalam penelitian ini, DBMS digunakan untuk menyimpan metadata file serta hasil ekstraksi teks dari file digital. DBMS juga mendukung integrasi dengan algoritma NLP, IR, dan TF-IDF untuk mendukung proses pencarian file yang efisien. Dengan peran ini, DBMS tidak hanya berfungsi sebagai tempat penyimpanan data, tetapi juga sebagai fondasi sistem manajemen file digital berbasis AI.

2.2.9 Mean Average Precision

Mean Average Precision (MAP) adalah salah satu metrik evaluasi yang paling umum digunakan dalam bidang *Information Retrieval* (IR), *Natural Language Processing* (NLP), serta sistem rekomendasi untuk menilai efektivitas pencarian atau penyajian informasi. MAP digunakan untuk mengukur seberapa baik suatu sistem dalam mengembalikan hasil yang relevan terhadap sebuah query pengguna, dengan memperhatikan urutan (ranking) dari hasil pencarian. Berbeda dengan metrik sederhana seperti *Precision* atau *Recall*, MAP memberikan perhatian khusus pada posisi dari dokumen relevan di dalam daftar hasil pencarian, sehingga mampu merepresentasikan kualitas ranking yang dihasilkan.

Pada dasarnya, MAP dihitung berdasarkan *Average Precision* (AP) dari setiap *query*. *Average Precision* sendiri merupakan rata-rata dari nilai presisi pada setiap titik di mana sebuah dokumen relevan ditemukan dalam daftar hasil pencarian. Setelah AP dihitung untuk masing-masing query, nilai MAP diperoleh dengan cara menghitung rata-rata dari semua AP pada seluruh *query* yang digunakan dalam pengujian. Dengan demikian, MAP memberikan gambaran umum mengenai performa sistem pencarian dalam skenario *multi-query*.

Secara formal, rumus MAP dapat dituliskan sebagai berikut:

$$MAP = \frac{1}{|Q|} \sum_{q=1}^{|Q|} AP(q)$$

dimana

- $|Q|$ = jumlah *query* dalam pengujian,
- $AP(q)$ = *Average Precision* dari *query* ke- q .

Sedangkan *average precision* dihitung dengan

$$AP = \frac{1}{R} \sum_{k=1}^N P(k) \cdot rel(k)$$

dimana

- R = jumlah dokumen relevan untuk sebuah *query*,
- N = jumlah total dokumen yang dikembalikan,
- $P(k)$ = nilai presisi pada posisi k ,
- $rel(k)$ = fungsi indikator yang bernilai 1 jika dokumen pada posisi k relevan, dan 0 jika tidak.

MAP memiliki beberapa keunggulan. Pertama, metrik ini memperhitungkan urutan dokumen, sehingga sistem yang mampu menempatkan dokumen relevan di posisi teratas akan memperoleh skor lebih tinggi. Kedua, MAP memberikan hasil yang konsisten pada skenario *multi-query*, berbeda dengan hanya menggunakan *Precision@k* atau *Recall* yang lebih terbatas. Ketiga, MAP sangat berguna untuk membandingkan dua sistem pencarian atau algoritma *ranking*, karena nilai yang dihasilkan bersifat *single value* dan mudah diinterpretasikan.

Namun demikian, MAP juga memiliki keterbatasan. Metrik ini mengasumsikan bahwa tingkat kepentingan atau relevansi setiap dokumen bersifat biner (relevan atau tidak relevan), padahal dalam praktiknya sebuah dokumen bisa memiliki tingkat relevansi yang berbeda-beda (*graded relevance*). Selain itu, MAP kurang ideal digunakan jika hanya ada sedikit *query* untuk evaluasi, karena rata-rata yang dihasilkan mungkin tidak cukup merepresentasikan performa sistem secara keseluruhan.

Dalam konteks sistem penelitian *Digital File Dashboard Management System with AI-Based Text Filtering and File Searching*, penggunaan MAP tentunya relevan untuk mengevaluasi kualitas pencarian dokumen. Misalnya, ketika pengguna mencari sebuah file dengan kata kunci tertentu, MAP dapat digunakan untuk mengukur apakah sistem mampu menampilkan file yang paling relevan di posisi awal hasil pencarian. Dengan demikian, MAP bukan hanya mengukur apakah file relevan ditemukan, tetapi juga apakah ia ditempatkan di posisi yang mudah diakses pengguna.

2.2.10 Normalized Discounted Cumulative Gain

Normalized Discounted Cumulative Gain (NDCG) merupakan salah satu metrik evaluasi yang banyak digunakan dalam bidang *information retrieval*, *machine learning*, serta *recommender system* untuk mengukur kualitas sistem perankingan. Tujuan utama dari NDCG adalah

menilai seberapa baik sebuah sistem mampu menampilkan hasil pencarian atau rekomendasi sesuai dengan relevansi yang diharapkan oleh pengguna. Tidak seperti metrik evaluasi sederhana seperti *precision* atau *recall*, NDCG mempertimbangkan posisi atau urutan dari hasil pencarian. Hal ini penting, sebab pada praktiknya, hasil yang lebih relevan dan muncul di peringkat atas memiliki nilai lebih tinggi dibanding hasil relevan yang ditempatkan pada posisi bawah.

Konsep NDCG berawal dari metrik Cumulative Gain (CG), yang menghitung jumlah total skor relevansi dari daftar hasil pencarian. CG kemudian dikembangkan menjadi Discounted Cumulative Gain (DCG), yang memasukkan faktor *discount* berdasarkan posisi peringkat. Faktor ini menggunakan formula logaritmik, sehingga relevansi pada posisi pertama dihargai lebih tinggi daripada relevansi yang sama pada posisi yang lebih rendah. Formula DCG umumnya dinyatakan sebagai berikut:

$$DCG_p = \sum_{i=1}^p \frac{2^{rel_i} - 1}{\log_2(i + 1)}$$

Dimana rel_i merupakan skor relevansi pada huruf ke- i , dan p adalah jumlah dokumen atau hasil yang dipertimbangkan. NDCG dapat berperan sebagai metrik evaluasi dalam menilai efektivitas algoritma pencarian file. Ketika pengguna melakukan pencarian, sistem tidak hanya dituntut untuk menemukan file yang relevan, tetapi juga untuk menampilkan file tersebut pada posisi atas daftar hasil pencarian. Misalnya, jika sistem berhasil menemukan dokumen yang sangat relevan tetapi menempatkannya pada peringkat bawah, hal ini tetap dinilai kurang optimal. Dengan menggunakan NDCG, peneliti dapat mengukur sejauh mana sistem yang dikembangkan mendekati urutan hasil pencarian yang ideal.

Lebih jauh, penggunaan NDCG juga memungkinkan evaluasi yang lebih realistis terhadap pengalaman pengguna (*user experience*). Pengguna biasanya hanya memperhatikan hasil pencarian pada beberapa posisi teratas, sehingga pengukuran berbasis posisi seperti NDCG dapat memberikan gambaran lebih jelas mengenai kegunaan praktis dari sistem. Dengan demikian, NDCG tidak hanya bermanfaat dari sisi akademis, tetapi juga memiliki dampak langsung terhadap kepuasan pengguna dalam interaksi sehari-hari dengan sistem pencarian file.

BAB 3 METODOLOGI

3.1 Metode yang digunakan

Dalam penelitian ini, pengembangan Sistem Informasi Manajemen Dasbor File Digital Dengan Filtering Text Dan Pencarian File Berbasis Kecerdasan Buatan menggunakan beberapa metode yaitu :

- ***Research & Development***

Penelitian ini menggunakan metode *Research and Development* (R&D) karena fokus utamanya adalah mengembangkan sebuah sistem manajemen dashboard digital yang dilengkapi dengan fitur pencarian dan penyaringan teks berbasis AI. Metode R&D memungkinkan peneliti untuk tidak hanya menghasilkan konsep teoritis, tetapi juga membangun prototipe nyata yang dapat diimplementasikan dan diuji.

R&D dipilih karena sesuai dengan tujuan penelitian, yaitu menciptakan sebuah produk perangkat lunak yang mampu meningkatkan efektivitas dalam mengelola file digital, mempercepat proses pencarian dokumen, serta memanfaatkan teknologi kecerdasan buatan untuk penyaringan teks. Dengan demikian, hasil penelitian tidak hanya memberikan kontribusi akademis, tetapi juga dapat digunakan secara praktis oleh pengguna akhir.

Secara umum, proses R&D mencakup beberapa tahapan, yaitu: (1) identifikasi potensi dan masalah, (2) pengumpulan data, (3) desain produk, (4) validasi desain, (5) revisi desain, (6) uji coba produk, (7) revisi produk, dan (8) implementasi. Dalam penelitian ini, beberapa tahapan akan disesuaikan dengan kebutuhan penelitian agar produk yang dihasilkan relevan dengan tujuan penelitian.

- ***Artificial Intelligence***

Untuk fitur pencarian berbasis AI, metodologi yang digunakan adalah penerapan *Information Retrieval* (IR) berbasis *Machine Learning*. Model pencarian akan dilatih dengan data teks yang sudah diproses melalui NLP, kemudian dibandingkan menggunakan algoritma vektor seperti *Cosine Similarity* atau *Word Embeddings* (contohnya Word2Vec atau BERT). Pendekatan ini memungkinkan sistem memahami konteks kata, bukan hanya kecocokan literal.

Tahap awal dilakukan dengan membangun indexing system pada semua dokumen yang masuk ke dalam dashboard. *Indexing* dilakukan menggunakan library seperti Elasticsearch atau Whoosh untuk mempercepat pencarian. Setelah indeks terbentuk, sistem akan menggunakan query user untuk mencocokkan hasil pencarian berdasarkan tingkat relevansi yang dihitung menggunakan bobot TF-IDF atau representasi embedding.

Agar lebih cerdas, metode semantic search berbasis transformer model (misalnya BERT) digunakan untuk memahami makna kalimat secara keseluruhan. Misalnya, ketika user mencari “laporan keuangan 2023”, sistem tidak hanya mencari kata literal “laporan

keuangan” tetapi juga bisa menampilkan dokumen dengan kata sinonim seperti “financial report 2023”.

Evaluasi dilakukan menggunakan metrik *Mean Average Precision* (MAP) atau *Normalized Discounted Cumulative Gain* (NDCG) untuk mengukur seberapa baik sistem menyajikan hasil pencarian yang relevan di peringkat teratas.

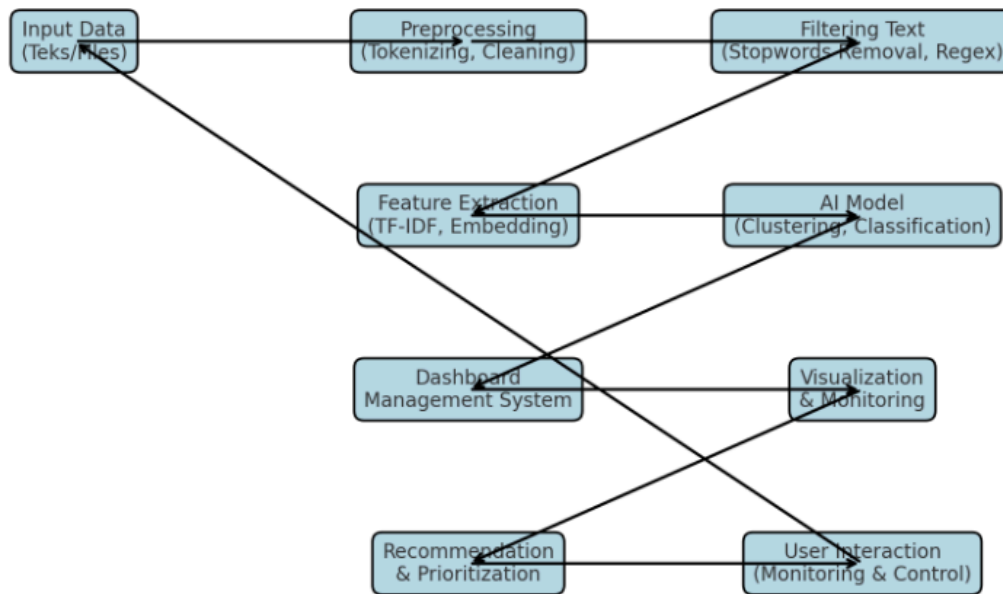
- ***Text filtering***

Fitur text filtering dikembangkan dengan tujuan menyaring konten file agar sesuai dengan kebutuhan pengguna. Proses pengembangannya diawali dengan tahap analisis kebutuhan, yaitu mengidentifikasi jenis file yang akan diproses (dokumen teks, PDF, atau dokumen office) dan menentukan kriteria penyaringan yang relevan, seperti kata kunci, kategori, atau topik. Selanjutnya dilakukan tahap data preprocessing yang mencakup pembersihan teks, normalisasi, penghapusan tanda baca, *stopwords removal*, serta *tokenization*. Metode ini penting agar data yang masuk ke sistem dapat diproses lebih akurat tanpa adanya gangguan dari elemen teks yang tidak relevan. Setelah itu, sistem text filtering diimplementasikan menggunakan algoritma pencocokan kata kunci (*keyword matching*) yang dikombinasikan dengan metode statistik seperti TF-IDF untuk memberikan penilaian bobot pada kata-kata dalam dokumen. Evaluasi dilakukan dengan menguji hasil filter terhadap dataset uji, membandingkan dokumen yang relevan dan tidak relevan, sehingga akurasi, presisi, dan recall dapat diukur sebagai indikator performa sistem.

Pada tahap filtering text, metodologi yang digunakan mengacu pada pendekatan **Natural Language Processing (NLP)** dengan proses **text preprocessing**. Proses ini melibatkan tahapan seperti tokenisasi, *stopword removal*, *stemming*, dan normalisasi teks untuk memastikan data teks menjadi lebih terstruktur sebelum dilakukan analisis. Tokenisasi digunakan untuk memecah kalimat menjadi kata-kata tunggal sehingga dapat diproses secara terpisah, sedangkan *stopword removal* bertujuan menghilangkan kata-kata umum yang tidak memiliki makna penting, seperti “dan”, “atau”, “yang”. Proses *stemming* dan *lemmatization* digunakan untuk menyamakan bentuk kata ke akar katanya, misalnya “berlari” menjadi “lari”, sehingga sistem dapat mengenali kata yang sama walaupun memiliki variasi bentuk.

Selanjutnya, dilakukan *feature extraction* dengan metode TF-IDF (*Term Frequency – Inverse Document Frequency*) untuk memberikan bobot pada kata yang sering muncul di dokumen tetapi jarang digunakan di dokumen lain, sehingga sistem dapat mengenali kata yang relevan. Hasil dari tahap ini adalah representasi teks dalam bentuk vektor numerik yang dapat digunakan dalam model pencarian atau analisis.

Metodologi ini juga akan mengintegrasikan filter berbasis *regex* (*regular expression*) untuk pencarian pola teks tertentu, misalnya pencarian berdasarkan nomor dokumen, tanggal, atau format standar lainnya. Dengan pendekatan ini, filtering text tidak hanya berbasis makna, tetapi juga mampu melakukan pencarian berbasis pola.



Gambar 3.1 Flowchart alur *filtering text*

Evaluasi dari tahap ini dilakukan dengan akurasi hasil filtering menggunakan dataset uji, misalnya dengan menghitung precision dan recall. Hal ini memastikan sistem filtering teks dapat memisahkan dokumen relevan dari dokumen yang tidak relevan dengan tingkat kesalahan minimal.

3.2 Bahan dan peralatan yang digunakan

3.2.1 Objek Penelitian

Objek penelitian ini adalah sistem manajemen digital file berbasis dashboard yang dilengkapi dengan fitur *AI-based text filtering* dan *file searching*. Sistem ini berfokus pada pengelolaan, pengorganisasian, serta pencarian dokumen digital secara lebih efisien dengan memanfaatkan metode kecerdasan buatan, khususnya pada pemrosesan bahasa alami (*Natural Language Processing*). Penelitian ini dilakukan untuk menganalisis, merancang, serta mengimplementasikan suatu sistem yang dapat membantu pengguna dalam mengelola dokumen, memfilter teks sesuai kebutuhan, serta melakukan pencarian file dengan hasil yang lebih relevan dan akurat. Dengan adanya objek penelitian ini, diharapkan dapat tercipta sebuah solusi teknologi yang mendukung kemudahan akses dan peningkatan efektivitas pengelolaan data digital dalam lingkup akademik maupun profesional.

3.2.2 Bahan dan peralatan yang digunakan

3.2.2.1 Hardware

Penulis menggunakan MacBook Pro M1 dengan prosesor Apple Silicon M1 yang memiliki arsitektur ARM, kelebihan utama perangkat ini adalah efisiensi daya yang tinggi, performa komputasi yang cepat, serta dukungan optimal untuk pengembangan perangkat lunak modern.

Penulis menggunakan smartphone Android sebagai perangkat uji coba aplikasi, dengan kelebihan fleksibilitas sistem operasi, dukungan luas terhadap berbagai aplikasi pengujian, serta kemudahan dalam melakukan integrasi dengan sistem berbasis web maupun notifikasi *real-time*.

3.2.2.2 Software

- **Visual Studio Code**

Peneliti menggunakan Visual Studio Code sebagai platform utama dalam proses pengembangan sistem manajemen berkas digital. VS Code dipilih karena memiliki beragam fitur yang mendukung produktivitas, seperti dukungan lintas bahasa pemrograman, integrasi dengan Git, serta ketersediaan ekstensi yang kaya untuk mendukung Laravel, JavaScript, maupun Python. Lingkungan pengkodean yang *user-friendly* membuat peneliti dapat bekerja lebih efektif dalam membangun fitur text filtering, file searching, dan integrasi dengan modul AI.

- **Laravel**

Peneliti menggunakan **Laravel** sebagai framework backend utama dalam membangun sistem manajemen dasbor. Laravel dipilih karena menawarkan arsitektur MVC yang jelas, sistem routing yang sederhana, serta fitur Eloquent ORM untuk pengelolaan basis data yang lebih efisien. Dengan Laravel, peneliti dapat mengembangkan aplikasi yang **scalable, maintainable, serta mudah diintegrasikan dengan modul AI dan search engine**. Penggunaan Laravel juga mempercepat proses pembuatan API yang dibutuhkan untuk menghubungkan frontend dengan backend..

- **ElasticSearch**

Peneliti menggunakan **ElasticSearch** sebagai search engine untuk mendukung proses pencarian informasi yang cepat dan relevan. ElasticSearch dipilih karena kemampuannya dalam melakukan **full-text search, indexing dokumen dalam jumlah besar, serta mendukung evaluasi pencarian menggunakan metrik seperti MAP dan nDCG**. Integrasi ElasticSearch dengan sistem memungkinkan pengguna untuk menemukan file berdasarkan kata kunci, relevansi, dan hasil pemfilteran teks berbasis AI.

- **Laravel Nova**

Peneliti menggunakan **Laravel Nova** sebagai *administration panel* untuk mengelola data dan sistem manajemen dasbor. Nova dipilih karena menyediakan antarmuka yang sederhana, elegan, dan mudah diintegrasikan langsung dengan Laravel. Dengan Nova, peneliti dapat dengan mudah melakukan **CRUD (Create, Read, Update, Delete)** terhadap file dan metadata, serta mengelola pengguna dalam sistem.

- **Jupyter Notebook**

Peneliti menggunakan **Jupyter Notebook** sebagai platform eksperimen dan evaluasi untuk modul AI dan NLP. Jupyter Notebook dipilih karena mendukung pemrograman interaktif dengan Python, sehingga memudahkan dalam **visualisasi hasil preprocessing teks, perhitungan TF-IDF, serta evaluasi model dengan MAP dan nDCG**. Notebook ini juga berperan penting dalam proses pengujian model sebelum diintegrasikan ke dalam sistem utama.

DAFTAR PUSTAKA

- Albahari, B., & Albahari, J. (2022). *C# 10 in a Nutshell: The Definitive Reference*. O'Reilly Media.
- Vilkov, Y. V., Chernysh, B. A., Kartamyshev, A. S., Golovenkin, E. N., & Poturemsky, I. V. (2019). The Document Management System as an Integration Basis for Forming a Unified Data Source of Information Support of a Product Life Cycle. *Journal of Physics: Conference Series*, 1399, 044099. DOI: 10.1088/1742-6596/1399/4/044099
- Baeza-Yates, R., & Ribeiro-Neto, B. (2011). *Modern Information Retrieval: The Concepts and Technology behind Search* (2nd ed.). Addison-Wesley.
- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4), 331–370. <https://doi.org/10.1023/A:1021240730564>
- Chapman, C. N., & McDonnell Feit, E. (2015). *R for Marketing Research and Analytics*. Springer.
- Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (4th ed.). SAGE Publications.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Tsani, A. B., Mahardika, F., & Santika, D. (2025). Pembuatan web dashboard interaktif untuk analisis data penjualan vending machine. *Modem: Jurnal Informatika dan Sains Teknologi*, 3(2), 51–61. <https://doi.org/10.62951/modem.v3i2.406>
- Hermawan, A. (2019). *Metode Penelitian Bisnis: Pendekatan Kuantitatif dan Kualitatif*. Kencana.
- ISO/IEC 9126. (2001). *Software Engineering — Product Quality*. International Organization for Standardization.
- Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4), 422–446. <https://doi.org/10.1145/582415.582418>
- Kumar, V., & Bansal, S. (2019). Evaluating information retrieval models using NDCG and MAP. *International Journal of Computer Applications*, 182(27), 1–6. <https://doi.org/10.5120/ijca2019918282>
- Laravel. (2023). *Laravel Documentation*. <https://laravel.com/docs>
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.

- Nielsen, J. (1994). *Usability Engineering*. Morgan Kaufmann.
- Pahl, C., & Hasselbring, W. (2001). An ontology for software component matching. *Proceedings of the Symposium on Software Reusability (SSR)*, 6–15.
- Pressman, R. S., & Maxim, B. R. (2014). *Software Engineering: A Practitioner's Approach* (8th ed.). McGraw-Hill Education.
- Rahman, A., & Nurdin, I. (2019). *Metodologi Penelitian: Kualitatif, Kuantitatif, dan Mixed Method*. Media Sahabat Cendekia.
- Royce, W. W. (1970). Managing the development of large software systems. *Proceedings of IEEE WESCON*, 1–9.
- Sommerville, I. (2015). *Software Engineering* (10th ed.). Pearson.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press.
- Tiwari, A., & Joshi, A. (2012). Comparative study of WebSocket and HTTP protocols. *International Journal of Computer Applications*, 56(13), 1–7.
- Zhang, Y., & Chen, X. (2020). Explainable recommendation: A survey and new perspectives. *Foundations and Trends in Information Retrieval*, 14(1), 1–101.
<https://doi.org/10.1561/15000000066>
- Wang, X. (2024). *The application of NLP in information retrieval*. *Applied and Computational Engineering*, 42, 290–297. DOI: 10.54254/2755-2721/42/20230795.

LAMPIRAN

BIODATA PENULIS