

PENERAPAN COSINE SIMILARITY DAN K-NEAREST NEIGHBOAR BERBASIS MACHINE LEARNING UNTUK Pencarian ILUSTRASI PADA KORAN INVESTOR DAILY

Renault Rinaldi, Devi Yunita

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang
Jl. Raya Puspipetek No.46, Buaran, Serpong, Kota Tangerang Selatan, Banten, Indonesia 15310,
renaultrinaldi22@gmail.com

ABSTRAK

Penelitian ini dilatarbelakangi oleh kebutuhan redaksi media cetak, khususnya *Investor Daily*, dalam mencari ilustrasi gambar yang relevan dengan konten berita secara cepat dan tepat. Permasalahan yang dihadapi adalah proses pencarian ilustrasi masih dilakukan secara manual berdasarkan penilaian subjektif, sehingga membutuhkan waktu lama dan kurang efisien. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sistem pencarian ilustrasi berbasis teks yang mampu memberikan hasil secara otomatis dan terukur. Metode yang digunakan adalah *Term Frequency-Inverse Document Frequency (TF-IDF)* untuk merepresentasikan teks *query* dan judul ilustrasi, *Cosine Similarity* untuk mengukur tingkat kemiripan, serta *K-Nearest Neighbor (K-NN)* untuk melakukan pemeringkatan hasil berdasarkan nilai kemiripan tertinggi. Sistem dikembangkan menggunakan bahasa pemrograman *Python* dengan *framework Flask* serta menggunakan *MySQL* sebagai basis data untuk menyimpan dan mengelola data ilustrasi. Hasil pengujian menunjukkan bahwa sistem mampu menampilkan ilustrasi yang relevan dengan *query*, seperti pada pencarian “kampus unpan”, di mana hasil dengan nilai *Cosine* tertinggi ditampilkan pada urutan teratas. Dengan demikian, sistem ini dapat meningkatkan efisiensi dan objektivitas dalam proses pencarian ilustrasi.

Kata kunci : , *K-Nearest Neighbor*, *Machine Learning*, *Information Retrieval*, *MySQL*

1. PENDAHULUAN

Latar Belakang Perkembangan teknologi informasi dan komunikasi telah memberikan dampak signifikan terhadap industri media, termasuk media cetak. Selain dituntut menyajikan informasi yang cepat dan akurat, media cetak juga harus mampu menghadirkan visualisasi berita yang menarik agar tetap relevan di tengah pesatnya perkembangan media digital. Ilustrasi berperan penting dalam memperjelas isi berita, menyederhanakan informasi kompleks, serta memperkuat pesan editorial.

Namun, proses pencarian ilustrasi pada redaksi *Investor Daily* masih dilakukan secara konvensional, sehingga membutuhkan waktu relatif lama, kurang efisien, dan cenderung subjektif. Kondisi ini menjadi kendala karena proses produksi koran berjalan di bawah tekanan waktu yang ketat. Oleh karena itu, diperlukan sistem pencarian ilustrasi otomatis berbasis teks yang mampu menemukan ilustrasi relevan secara cepat dan akurat dengan pendekatan *Information Retrieval* dan *Machine Learning*.

Identifikasi Masalah, berdasarkan latar belakang tersebut, permasalahan yang dihadapi dalam penelitian ini meliputi proses pencarian ilustrasi yang masih konvensional dan memakan waktu, belum tersedianya sistem otomatis untuk membantu redaksi menemukan ilustrasi yang relevan, serta belum diterapkannya metode *machine learning* berbasis teks untuk meningkatkan efektivitas pencarian ilustrasi.

Rumusan Masalah Rumusan masalah dalam penelitian ini adalah bagaimana merancang dan mengimplementasikan sistem pencarian ilustrasi

otomatis berbasis teks menggunakan algoritma *Cosine Similarity* dan *K-Nearest Neighbor*, bagaimana kemampuan *Cosine Similarity* dalam mengukur tingkat kemiripan semantik antara teks berita dan anotasi ilustrasi, serta bagaimana penerapan *K-Nearest Neighbor* dapat meningkatkan relevansi dan akurasi hasil pencarian ilustrasi.

Tujuan Penelitian Penelitian ini bertujuan untuk merancang dan mengembangkan sistem pencarian ilustrasi otomatis berbasis teks, mengimplementasikan *Cosine Similarity* untuk mengukur kemiripan semantik, menerapkan *K-Nearest Neighbor* dalam pemeringkatan hasil pencarian, serta menguji efektivitas sistem pada lingkungan kerja redaksi *Investor Daily*.

Manfaat Penelitian Manfaat penelitian ini meliputi peningkatan pemahaman penulis dalam penerapan *machine learning* pada sistem temu kembali informasi berbasis teks, kontribusi akademik bagi Universitas Pamulang sebagai referensi ilmiah di bidang *Information Retrieval*, serta manfaat praktis bagi *Investor Daily* dalam mempercepat proses pencarian ilustrasi dan meningkatkan kualitas visual penyajian berita.

Batasan Masalah Ruang lingkup penelitian dibatasi pada pengolahan anotasi teks ilustrasi tanpa melibatkan pemrosesan citra visual. Dataset bersumber dari arsip ilustrasi *Investor Daily* dengan jumlah terbatas, menggunakan metode TF-IDF, *Cosine Similarity*, dan *K-Nearest Neighbor*. Sistem dibangun menggunakan *Python*, *framework Flask*, dan

basis data MySQL, dengan evaluasi kinerja yang difokuskan pada metrik akurasi pencarian.

2. LANDASAN TEORI

2.1. Penelitian Terkait

Penelitian ini mengacu pada sejumlah studi sebelumnya yang membahas penerapan *Information Retrieval* berbasis teks menggunakan metode *TF-IDF*, *Cosine Similarity*, dan *K-Nearest Neighbor (K-NN)*.

Beberapa penelitian menunjukkan bahwa metode *Term Frequency-Inverse Document Frequency (TF-IDF)* efektif dalam merepresentasikan dokumen teks ke dalam bentuk numerik sehingga dapat digunakan dalam proses pencarian informasi (Salton & Buckley, 1988). Metode *Cosine Similarity* banyak digunakan untuk mengukur tingkat kemiripan antar dokumen berdasarkan sudut antar vektor, yang terbukti memberikan hasil yang akurat dan efisien dalam sistem pencarian berbasis teks.

Selain itu, algoritma *K-Nearest Neighbor (K-NN)* digunakan untuk melakukan klasifikasi atau pemeringkatan berdasarkan kedekatan jarak antar data, yang mampu meningkatkan relevansi hasil pencarian.

Berbeda dengan penelitian sebelumnya, penelitian ini mengintegrasikan metode *Cosine Similarity* dan *K-NN* secara spesifik untuk pencarian ilustrasi pendukung grafis pada media cetak, khususnya di lingkungan redaksi *Investor Daily*, yang memiliki keterbatasan waktu dan tingkat subjektivitas tinggi dalam pemilihan ilustrasi.

2.2. Kerangka Berpikir

Permasalahan utama dalam penelitian ini adalah proses pencarian ilustrasi yang masih dilakukan secara konvensional, sehingga memerlukan waktu yang lama dan bersifat subjektif. Untuk mengatasi permasalahan tersebut, penelitian ini mengusulkan sistem pencarian ilustrasi otomatis berbasis teks dengan pendekatan *Information Retrieval* dan *Machine Learning*.

Alur sistem dimulai dari tahap *preprocessing* teks berita dan anotasi ilustrasi menggunakan teknik *Natural Language Processing (NLP)*, seperti *case folding*, tokenisasi, dan *stopword removal*. Teks yang telah diproses kemudian direpresentasikan ke dalam bentuk vektor numerik menggunakan metode *TF-IDF*.

Selanjutnya, metode *Cosine Similarity* digunakan untuk menghitung tingkat kemiripan antara teks berita dan deskripsi ilustrasi. Nilai kemiripan tersebut kemudian diproses menggunakan algoritma *K-Nearest Neighbor (K-NN)* untuk menentukan ilustrasi yang paling relevan berdasarkan nilai kedekatan tertinggi.

Integrasi metode tersebut diharapkan mampu meningkatkan kecepatan pencarian, mengurangi subjektivitas dalam pemilihan ilustrasi, serta meningkatkan kualitas visual penyajian berita pada media cetak.

2.3. Information Retrieval

Merupakan bidang ilmu yang mempelajari proses pencarian informasi yang relevan dari kumpulan data dalam jumlah besar berdasarkan kebutuhan pengguna (*query*). Sistem *IR* bekerja dengan mencocokkan representasi *query* dengan dokumen yang tersimpan, kemudian menghasilkan peringkat dokumen berdasarkan tingkat relevansinya.

Dalam implementasinya, *IR* melibatkan beberapa tahapan utama seperti *preprocessing* teks, representasi dokumen, serta pengukuran kemiripan. *IR* digunakan untuk mencocokkan teks berita dengan anotasi ilustrasi, sehingga sistem dapat memberikan rekomendasi gambar yang sesuai secara otomatis.

2.4. Natural Language Processing (NLP)

Merupakan cabang dari kecerdasan buatan yang berfokus pada pemrosesan bahasa alami manusia agar dapat dipahami oleh komputer. *NLP* memungkinkan sistem untuk melakukan analisis teks melalui berbagai tahapan *preprocessing* seperti *case folding*, tokenisasi, *stopword removal*, dan *stemming*.

Tahapan *preprocessing* sangat penting dalam meningkatkan kualitas data teks karena mampu mengurangi noise serta menyederhanakan variasi kata. Dalam penelitian ini, *NLP* digunakan untuk menstandarkan teks berita dan anotasi ilustrasi sebelum dilakukan proses pembobotan dan perhitungan kemiripan.

2.5. Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF merupakan metode pembobotan kata yang digunakan untuk merepresentasikan dokumen dalam bentuk vektor numerik. Metode ini menggabungkan dua komponen utama, yaitu *Term Frequency (TF)* yang mengukur frekuensi kemunculan kata dalam dokumen, dan *Inverse Document Frequency (IDF)* yang mengukur tingkat keunikan kata dalam keseluruhan koleksi.

Kata yang sering muncul dalam suatu dokumen namun jarang muncul pada dokumen lain akan memiliki bobot tinggi, sehingga dianggap lebih penting. Dalam penelitian ini, *TF-IDF* digunakan sebagai dasar dalam membentuk representasi vektor teks yang akan digunakan dalam perhitungan kemiripan.

2.6. Cosine Similarity

Cosine Similarity merupakan metode yang digunakan untuk mengukur tingkat kemiripan antara dua vektor dengan menghitung sudut di antara keduanya. Nilai *cosine similarity* berada pada rentang 0 hingga 1, di mana nilai yang mendekati 1 menunjukkan tingkat kemiripan yang tinggi.

Metode ini banyak digunakan dalam sistem pencarian informasi karena tidak dipengaruhi oleh panjang dokumen, melainkan fokus pada arah vektor. Dalam penelitian ini, *cosine similarity* digunakan untuk menghitung kedekatan antara *query* pengguna

dengan dokumen ilustrasi yang telah direpresentasikan menggunakan TF-IDF.

2.7. K-Nearest Neighbor (K-NN)

K-Nearest Neighbor (K-NN) merupakan algoritma *machine learning* berbasis jarak yang digunakan untuk klasifikasi maupun pemeringkatan data berdasarkan kedekatan terhadap data lain. Algoritma ini bekerja dengan mencari sejumlah K tetangga terdekat dari suatu data, kemudian menentukan hasil berdasarkan mayoritas atau nilai kedekatan.

Dalam konteks penelitian ini, K-NN digunakan untuk melakukan pemeringkatan hasil pencarian ilustrasi berdasarkan nilai kemiripan tertinggi. Dengan demikian, sistem tidak hanya menampilkan hasil yang relevan, tetapi juga mengurutkannya berdasarkan tingkat kedekatan dengan query.

2.8. Basis Data dan Pemodelan Sistem

Basis data merupakan komponen penting dalam sistem informasi yang berfungsi untuk menyimpan, mengelola, dan mengakses data secara terstruktur. Pada penelitian ini, digunakan MySQL sebagai sistem manajemen basis data relasional karena kemampuannya dalam menangani data secara efisien dan mendukung integritas.

Pemodelan sistem dilakukan menggunakan Unified Modeling Language (UML), yang merupakan bahasa standar untuk memvisualisasikan, merancang, dan mendokumentasikan sistem perangkat lunak. UML membantu dalam menggambarkan struktur sistem serta interaksi antar komponen secara sistematis.

2.9. Entity Relationship Diagram (ERD)

Entity Relationship Diagram (ERD) digunakan untuk memodelkan struktur logis basis data melalui representasi entitas, atribut, dan relasi antar entitas. ERD membantu dalam merancang database yang konsisten serta meminimalkan redundansi data.

Dalam penelitian ini, ERD digunakan sebagai acuan dalam membangun struktur tabel yang mencakup data ilustrasi, anotasi teks, serta relasi antar data yang mendukung proses pencarian ilustrasi.

2.10. Unified Modeling Language (UML)

Unified Modeling Language (UML) merupakan bahasa pemodelan visual yang digunakan untuk mendeskripsikan sistem perangkat lunak secara terstruktur. UML terdiri dari berbagai jenis diagram seperti *use case diagram*, *activity diagram*, *sequence diagram*, dan *class diagram* yang masing-masing memiliki fungsi berbeda dalam menggambarkan sistem.

Dalam penelitian ini, UML digunakan untuk memodelkan alur sistem pencarian ilustrasi, mulai dari

interaksi pengguna hingga proses pengolahan data dan penyajian hasil, sehingga memudahkan dalam pengembangan dan implementasi sistem.

3. ANALISIS DAN PERANCANGAN SISTEM

3.1. Analisis Kebutuhan

Kebutuhan Fungsional adalah Analisis terhadap proses bisnis, tujuan penelitian, serta mekanisme kerja sistem pencarian ilustrasi berbasis teks. Kebutuhan ini menggambarkan fungsi utama yang harus disediakan sistem agar dapat beroperasi secara optimal dan memenuhi kebutuhan pengguna.

Kebutuhan Non-Fungsional adalah Kebutuhan ini mencakup aspek-aspek kualitas yang harus dipenuhi agar sistem dapat berjalan dengan baik, mulai dari performa, akurasi, kemudahan penggunaan, keamanan, hingga portabilitas. Seluruh aspek tersebut menjadi acuan dalam memastikan sistem dapat digunakan secara optimal oleh pengguna.

3.2. Analisis Sistem Berjalan

Sistem yang berjalan saat ini dalam proses pemilihan ilustrasi berita di lingkungan redaksi masih dilakukan secara *manual dan konvensional*. Redaktur atau tim desain selama ini harus mencari ilustrasi secara visual berdasarkan ingatan, folder penyimpanan lokal, atau penamaan file yang terbatas. Proses ini umumnya dilakukan dengan membuka satu per satu folder gambar tanpa adanya mekanisme pencarian berbasis isi teks berita.

3.3. Rumus TF-IDF

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

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

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

Rumus Cosine Similarty

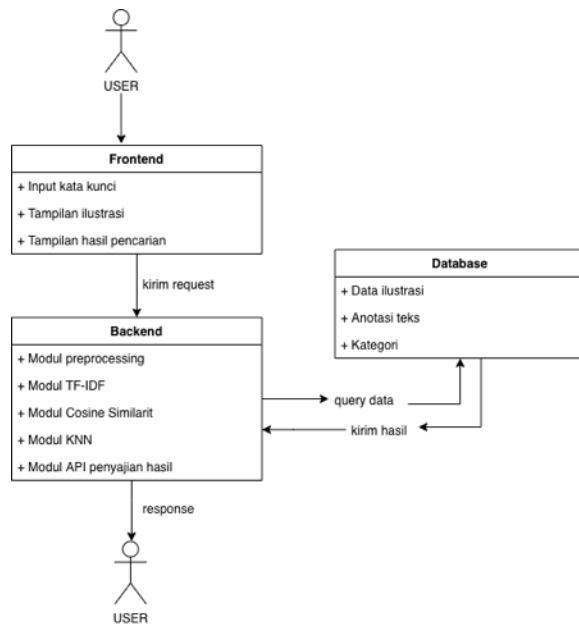
$$CosSim(A, B) = \frac{\sum A_i B_i}{\sqrt{\sum A_i^2} \cdot \sqrt{\sum B_i^2}}$$

Rumus K-Nearest Neighbor (K-NN).

$$d(x, y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2}$$

3.4. Arsitektur Sistem

Pada penelitian ini dirancang menggunakan model *client-server* yang terdiri dari tiga komponen utama, yaitu *Frontend*, *Backend*, dan *Database*, dengan pengguna (*user*) sebagai aktor yang berinteraksi langsung dengan sistem.



Gambar 1. Arsitektur Sistem

3.5. Perancangan Layar (User Interface)

Perancangan layar (user interface) merupakan tahap penting dalam pengembangan sistem karena berfungsi sebagai media interaksi antara pengguna dan sistem. Perancangan layar pada sistem pencarian ilustrasi ini bertujuan untuk menghasilkan antarmuka yang mudah digunakan, informatif, dan mampu mendukung proses pencarian ilustrasi berbasis teks secara efektif.

Perancangan Layar *Login* Layar login dirancang sebagai halaman awal yang digunakan untuk mengakses sistem. Pengguna diwajibkan melakukan autentikasi dengan memasukkan username dan password yang telah terdaftar. Komponen pada layar *login*, untuk membatasi akses sistem hanya kepada pengguna yang memiliki hak akses, serta menjaga keamanan data ilustrasi.



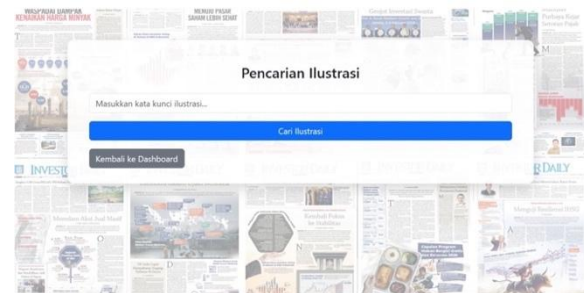
Gambar 2. Tampilan Halaman Login

Perancangan Layar Beranda (*Home*), Layar beranda merupakan halaman utama yang ditampilkan setelah pengguna berhasil *login*. Pada layar ini, pengguna dapat langsung melakukan pencarian ilustrasi tanpa harus melalui menu yang kompleks. Komponen pada layar beranda

Perancangan layar beranda difokuskan pada kemudahan akses fitur pencarian agar pengguna dapat

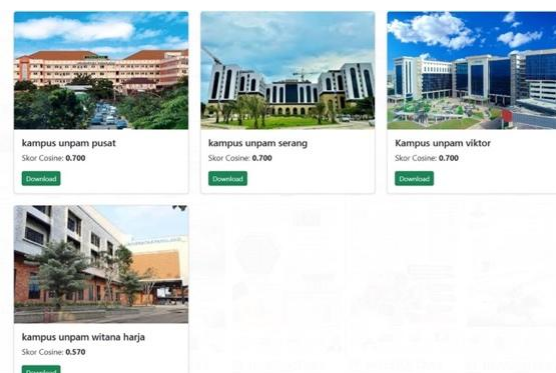
langsung memasukkan kata kunci sesuai kebutuhan ilustrasi.

Kemudian, Perancangan Layar *Input Kata Kunci* Layar input kata kunci dirancang untuk menerima masukan teks dari pengguna berupa kata kunci pencarian ilustrasi. Kata kunci ini biasanya berasal dari judul atau isi berita yang sedang dikerjakan. Komponen pada layar input kata kunci. Layar ini menjadi titik awal proses pengolahan teks yang selanjutnya akan melalui tahapan preprocessing, pembobotan TF-IDF, dan perhitungan kemiripan.



Gambar 3. Tampilan Input Kata Kunci

Kemudian, Perancangan Layar Hasil Pencarian Ilustrasi Layar hasil pencarian dirancang untuk menampilkan ilustrasi yang paling relevan berdasarkan kata kunci yang dimasukkan pengguna. Hasil pencarian ditampilkan dalam bentuk daftar atau grid agar mudah dilihat. Komponen pada layar hasil pencarian. Ilustrasi ditampilkan berdasarkan urutan tingkat kemiripan tertinggi ke terendah, sehingga pengguna dapat langsung memilih ilustrasi yang paling sesuai.



Gambar 4. Tampilan Hasil Pencarian

Kemudian, Perancangan Layar *Detail Ilustrasi*, layar detail ilustrasi digunakan untuk menampilkan informasi lebih lengkap mengenai ilustrasi yang dipilih. Pada layar ini, pengguna dapat melihat gambar ilustrasi dalam ukuran lebih besar serta deskripsi lengkapnya. Komponen pada layar detail ilustrasi layar ini membantu pengguna dalam memastikan kesesuaian ilustrasi sebelum digunakan dalam layout berita.

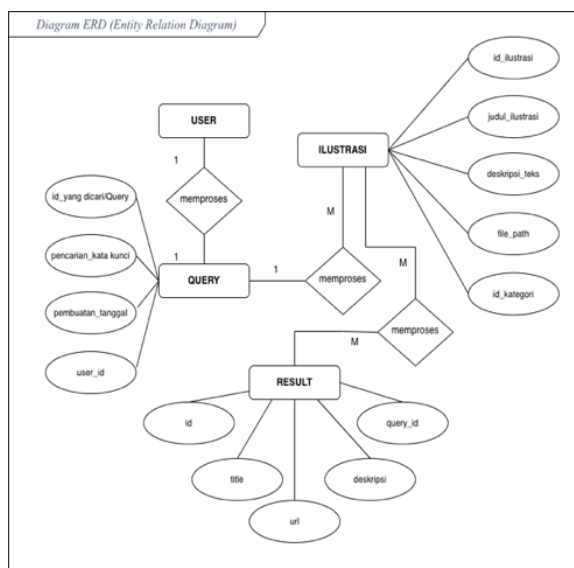
Dilanjutkan dengan perancangan Layar *Logout* yang dirancang sebagai mekanisme untuk mengakhiri

sesi penggunaan sistem. Ketika pengguna memilih menu logout, sistem akan menghapus sesi pengguna dan mengarahkan kembali ke layar *login*. Tujuan dari perancangan layar *logout* adalah untuk menjaga keamanan sistem dan mencegah akses tidak sah apabila sistem ditinggalkan dalam keadaan aktif.

3.6. Perancangan Basis Data

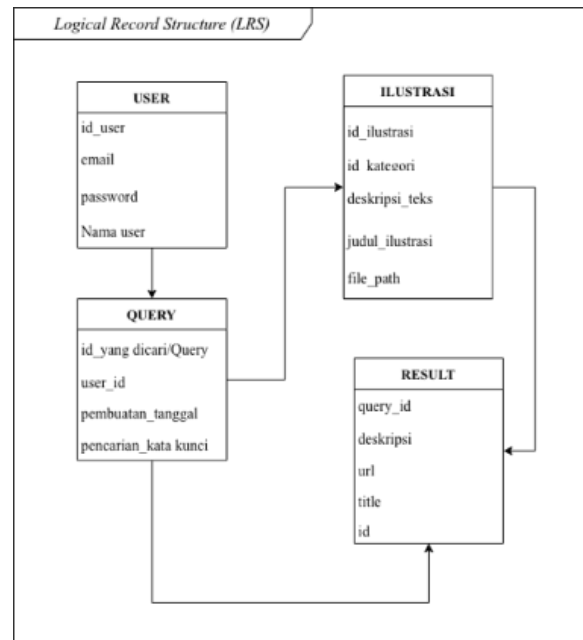
Perancangan basis data merupakan tahapan penting dalam pengembangan sistem informasi karena berfungsi sebagai fondasi utama dalam proses penyimpanan, pengelolaan, dan pengambilan data secara terstruktur. Pada penelitian ini, basis data dirancang untuk mendukung sistem pencarian ilustrasi berbasis teks dengan pendekatan *Information Retrieval* menggunakan metode *TF-IDF*, *Cosine Similarity*, dan *K-Nearest Neighbor (K-NN)*.

Entity Relationship Diagram (ERD) digunakan untuk menginterpretasikan, menentukan, dan mendokumentasikan kebutuhan-kebutuhan untuk sistem pemrosesan pada database. Adapun database pada sistem pencarian ilustrasi untuk pendukung tulisan pada koran *Investor Daily* menggunakan metode *Cosine Similarity* dan *K-NN* digambarkan dalam bentuk *ERD* sebagai berikut:



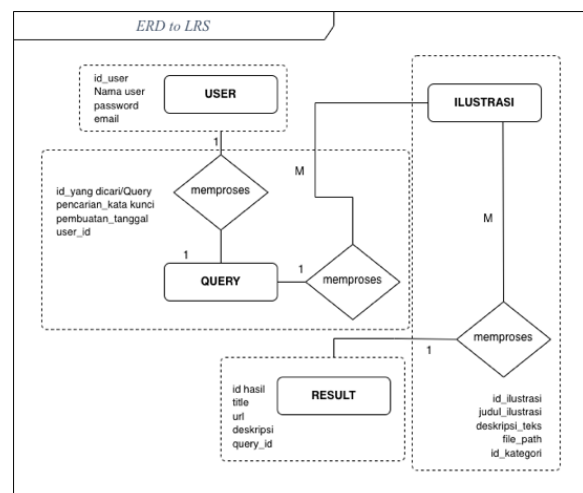
Gambar 5. Entity Relationship Diagram (ERD)

Kemudian Transformasi *ERD* ke *Logical Record Structure (LRS)* Proses transformasi ini bertujuan untuk menerjemahkan setiap entitas pada ERD menjadi struktur tabel logis, menentukan *primary key*, *foreign key*, serta hubungan antar tabel secara eksplisit. Dengan adanya LRS, pengembang dapat memastikan bahwa struktur data yang dirancang telah konsisten, terhindar dari redundansi, serta mendukung kebutuhan sistem, khususnya dalam proses pencarian ilustrasi berbasis teks menggunakan metode *TF-IDF*, *Cosine Similarity*, dan *K-Nearest Neighbor (K-NN)*. Adapun transformasi ERD di atas ke dalam bentuk LRS dapat dilihat pada gambar berikut:



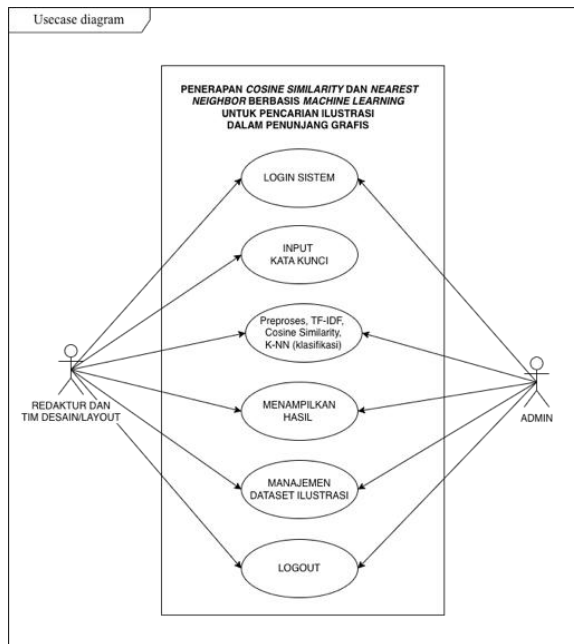
Gambar 6. Transformasi Entity Relationship Diagram (ERD) ke dalam Logical Record Structure (LRS)

Kemudian dilanjutkan pada *Logical Record Structure (LRS)*. Setelah *ERD* ditransformasikan maka akan menghasilkan *Logical Record Structure (LRS)* sebagai berikut:



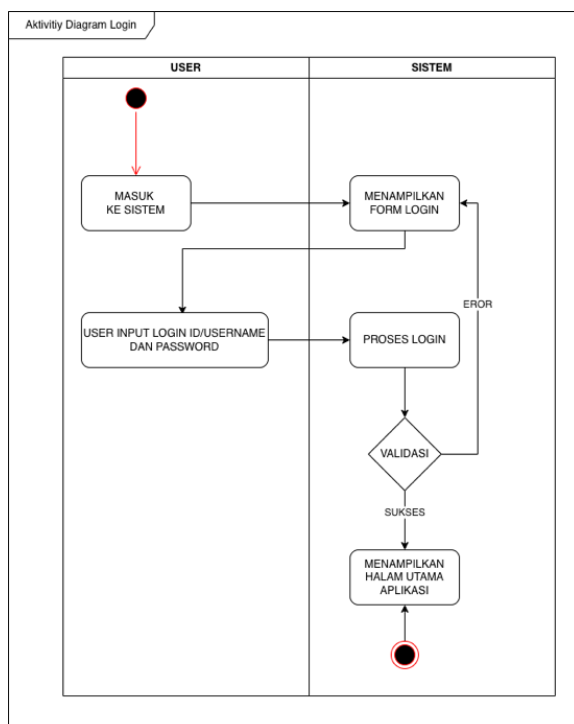
Gambar 7. menghasilkan Logical Record Structure (LRS)

Kemudian, *Use case diagram* dibuat untuk menjelaskan apa yang dilakukan oleh sistem dan *actor* sebagai pengguna yang berhubungan dengan proses pada sistem yang akan dibuat. Pada sistem pencarian ilustrasi untuk penunjang grafis di media koran *Investor daily*, menggunakan metode *Cosine Similarity* dan *K-NN*.



Gambar 8. Use case diagram

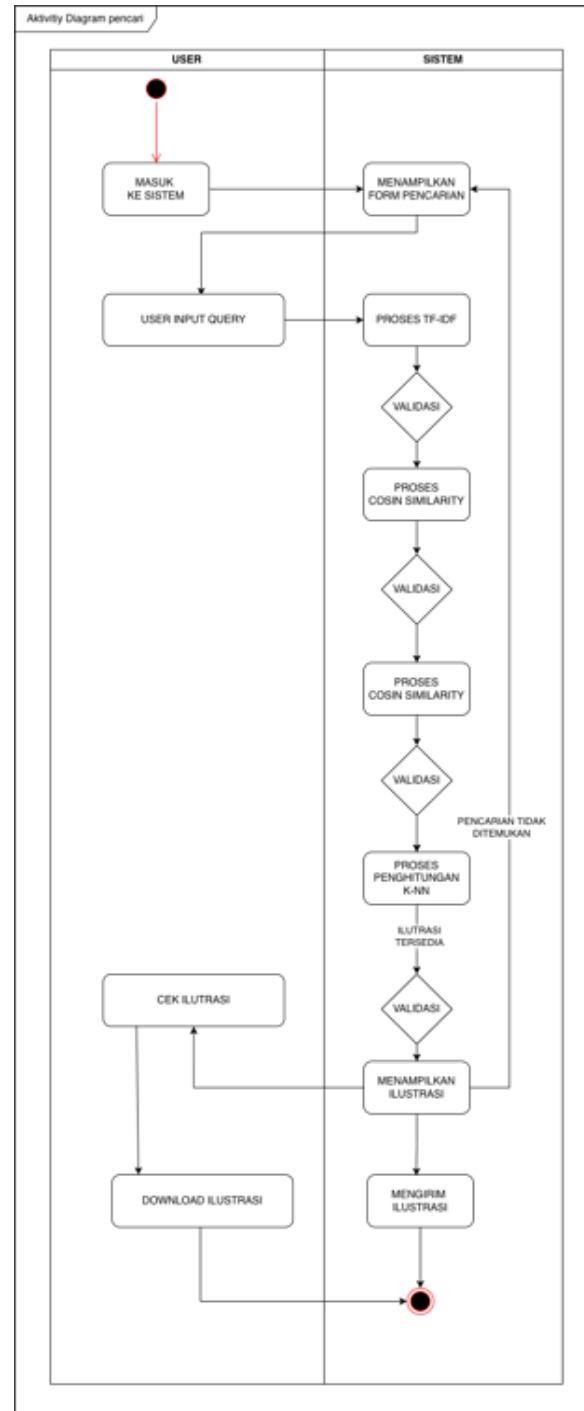
Kemudian masuk ke diagram *aktiviti login*, dibuat untuk menjelaskan apa yang dilakukan oleh sistem dan *actor* sebagai pengguna yang berhubungan dengan proses pada sistem yang akan dibuat. Pada sistem pencarian ilustrasi untuk penunjang grafis di media koran *Investor daily*, menggunakan metode *Cosine Similarity* dan *K-NN*, dari mulai *login*, dapat dilihat pada gambar berikut:



Gambar 9. Activity Diagram Login

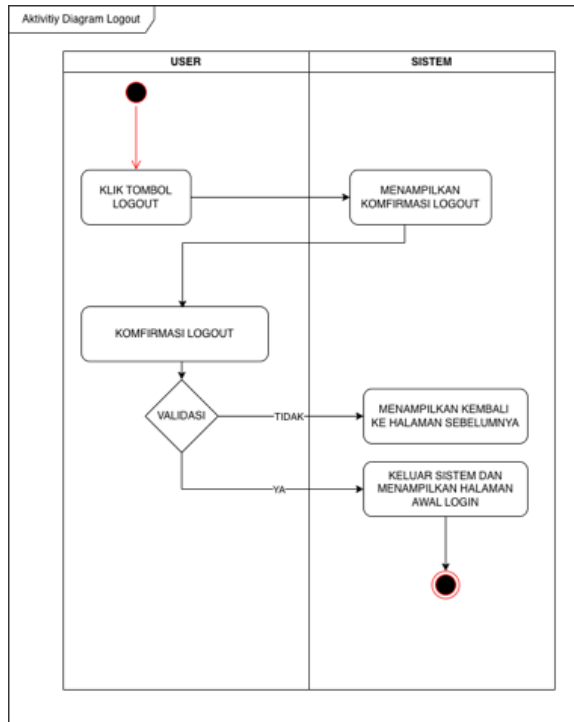
Kemudain *Aktiviti diagram pencarian Diagram aktiviti diagram login*, dibuat untuk menjelaskan apa yang dilakukan oleh sistem dan *actor* sebagai

pengguna yang berhubungan dengan proses pada sistem yang akan dibuat. Pada sistem pencarian ilustrasi untuk penunjang grafis di media koran *Investor daily*, menggunakan metode *Cosine Similarity* dan *K-NN*, dari mulai *login*, dapat dilihat pada gambar:



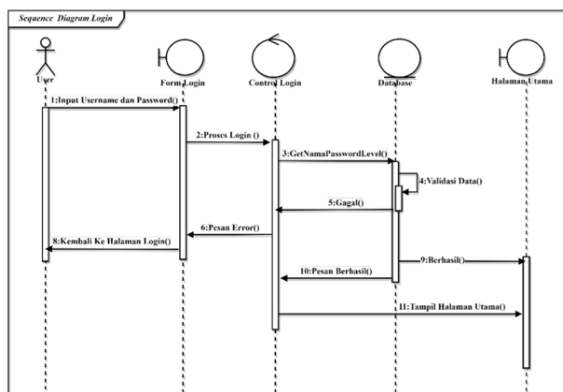
Gambar 10. Activity Diagram Pecarian

Kemudian dilanjut pada Aktiviti diagram *logout* Diagram aktiviti diagram *login*, dibuat untuk menjelaskan apa yang dilakukan oleh sistem dan *actor* sebagai pengguna yang berhubungan dengan proses pada sistem yang akan dibuat. Pada sistem pencarian ilustrasi untuk penunjang grafis di media koran *Investor daily*, menggunakan metode *Cosine Similarity* dan *K-NN*, dari mulai *login*, dapat dilihat pada gambar

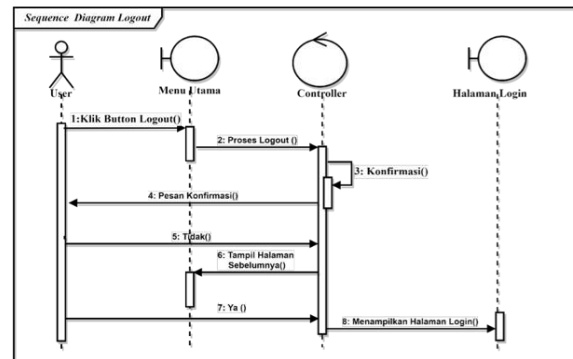


Gambar 11. Sequence Diagram Logout

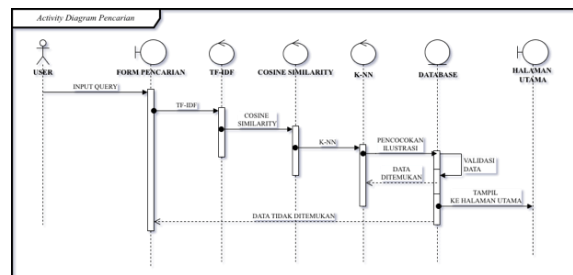
Sequence diagram dibuat untuk menunjukkan aliran fungsional dalam *use case diagram*. Adapun *sequence diagram* pada sistem yang akan dibuat adalah *sequence diagram login*, *sequence diagram pencarian* dan *sequence diagram logout*.



Gambar 12. Sequence Diagram Login

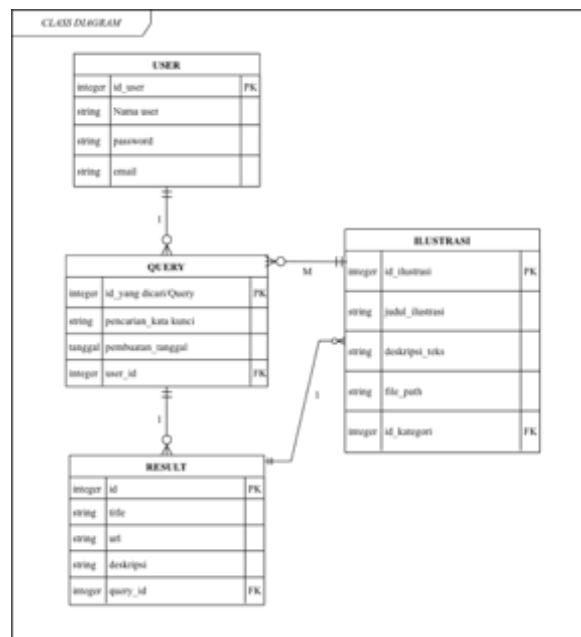


Gambar 13. Sequence Diagram Login



Gambar 14. Sequence Diagram Pencarian

Kemudian, lanjut ke *Class Diagram* yang digunakan untuk menampilkan kelas-kelas didalam model desain suatu sistem dengan memperlihatkan aturan-aturan dan tanggung jawab entitas. Adapun *class diagram* dari penerapan *Cosine Similarity* dan *nearest neighbor* berbasis *machine learning* untuk pencarian ilustrasi dalam penunjang grafis pada media koran *Investor Daily* berbasis web adalah sebagai berikut:



Gambar 15. Class Ciagram

4. HASIL DAN PENGUJIAN

4.1. Implementasi Sistem

Implementasi dan pengujian dari sistem pencarian ilustrasi berbasis teks yang telah dirancang pada Bab III. Dimana implementasi sistem dilakukan untuk merealisasikan perancangan yang telah dibuat agar dapat digunakan secara nyata oleh pengguna, khususnya redaktur dan tim desain/layout pada media cetak *Investor Daily*.

Sistem dikembangkan dalam bentuk aplikasi berbasis web menggunakan bahasa pemrograman Python dengan framework Flask. Pemilihan Flask didasarkan pada sifatnya yang ringan, fleksibel, dan mudah diintegrasikan dengan library *Machine Learning* serta Natural Language Processing (NLP). Basis data yang digunakan adalah *MySQL* untuk menyimpan data ilustrasi, kategori, pengguna, serta riwayat pencarian.

4.2. Lingkungan Implementasi

Perangkat Keras (*Hardware*) Perangkat keras ini berfungsi untuk mendukung proses perancangan, implementasi, serta pengujian aplikasi agar dapat berjalan dengan baik dan optimal.

Tabel 1. Spesifikasi Perangkat Keras (Hardware)

No	Nama	Spesifikasi
1	Merk	Lenovo Laptop
2	Prosesor	Intel Core i5 / setara
3	RAM	8 GB
4	Penyimpanan	SSD 256 GB
5	Perangkat Tampilan	Laptop/PC

Perangkat Lunak (*Software*) Perangkat lunak yang digunakan meliputi:

Tabel 2. Perangkat Lunak (Software)

No	Nama	Spesifikasi
1	Sistem Operasi	Windows 10
2	Bahasa Pemrograman	Python 3.x
3	Framework Web	Flask
4	Basis Data	MySQL
5	Text Editor / IDE	Visual Studio Code
6	Library Pendukung	- Sastrawi (stemming Bahasa Indonesia) - NLTK (tokenisasi dan stopword removal) - Scikit-learn (TF-IDF, Cosine Similarity, dan K-NN) - Draw.io - Pandas dan NumPy (pengolahan data)

5. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis, perancangan, implementasi, serta pengujian sistem yang telah dilakukan, maka dapat disimpulkan sebagai berikut:

Sistem pencarian ilustrasi berbasis teks yang dikembangkan mampu membantu redaktur dan tim desain dalam menemukan ilustrasi yang relevan

dengan isi berita secara lebih cepat, akurat, dan efisien dibandingkan dengan metode pencarian manual.

Penerapan tahapan *text preprocessing* yang dipadukan dengan metode TF-IDF dan *Cosine Similarity* berhasil merepresentasikan teks ke dalam bentuk vektor numerik, sehingga memungkinkan pengukuran tingkat kemiripan antara query pencarian.

Algoritma *K-Nearest Neighbor (K-NN)* berperan penting dalam proses klasifikasi dan pemeringkatan hasil pencarian dengan menghitung jarak kedekatan antar vektor dan menentukan ilustrasi berdasarkan sejumlah tetangga terdekat. Nilai tertinggi yang dihasilkan oleh *K-NN* berada di kisaran 0,89.

Beberapa saran yang dapat dijadikan bahan pengembangan sistem di masa mendatang, antara lain: Sistem pencarian ilustrasi ini dapat dikembangkan lebih lanjut guna meningkatkan akurasi hasil pencarian.

Pengembangan fitur pencarian dapat ditingkatkan dengan menambahkan analisis citra (*image-based search*) sehingga pencarian tidak hanya bergantung pada deskripsi teks, tetapi juga pada kemiripan visual ilustrasi.

Dataset ilustrasi yang digunakan sebaiknya diperbanyak dan diperbarui secara berkala agar sistem dapat memberikan hasil pencarian yang lebih variatif dan relevan.

Sistem dapat dikembangkan ke dalam platform lain seperti aplikasi mobile agar lebih fleksibel dan mudah digunakan oleh pengguna.

Diperlukan pengujian sistem dengan jumlah pengguna dan data yang lebih besar untuk memperoleh hasil evaluasi yang lebih komprehensif.

DAFTAR PUSTAKA

- [1] Djunaidi, A., Ningrum, D. R., Kusuma, A., & Saradiba, R. (2024). Optimasi pencarian data obat menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)* dan *Cosine Similarity*. *Jurnal Teknologi Informasi*, 9(2), 112–120.
- [2] Hendini. (2020). Pemodelan sistem menggunakan *Unified Modeling Language (UML)*. Dalam O. Surbakti (Ed.), *Rekayasa Perangkat Lunak*. Jakarta: Andi.
- [3] Hermiati, R., Asnawati, & Kanedi, I. (2021). *Sistem manajemen basis data relasional*. Bandung: Informatika.
- [4] Pramudito, A., & Handayani, R. (2022). Klasifikasi dokumen berita online menggunakan algoritma *K-Nearest Neighbor*. *Jurnal Ilmu Komputer*, 7(1), 45–53.
- [5] Pratama, I. P. A. E., & Saparingga, R. (2021). *Pemodelan perangkat lunak menggunakan Unified Modeling Language (UML)*. Yogyakarta: Andi.
- [6] Rasyid, A., & Ningsih, S. (2024). *Information retrieval* berbasis teks pada sistem pencarian dokumen. *Jurnal Sistem Informasi*, 10(1), 1–10.

- [7] Santoti, Jocelyn, & Irsyad. (2024). Analisis kemiripan produk halal menggunakan TF-IDF dan *Cosine Similarity*. *Jurnal Teknologi Informasi dan Komputer*, 11(2), 89–98.
- [8] Sari, M., Hartati, S., & Prasetyo, E. (2023). Sistem rekomendasi artikel ilmiah berbasis kesamaan konten menggunakan algoritma *K-Nearest Neighbor*. *Jurnal Informatika*, 8(2), 134–142.
- [9] Siregar, R., & Siagian, H. (2021). *Perancangan basis data dengan Entity Relationship Diagram (ERD)*. Yogyakarta: Deepublish.
- [10] Sukamto, R. A., & Shalahuddin, M. (2021). *Rekayasa perangkat lunak terstruktur dan berorientasi objek*. Bandung: Informatika.
- [11] Surbakti, A. (2020). *Konsep dan penerapan sistem informasi*. Jakarta: Gramedia.
- [12] Tabrani, M., Suhardi, & Priyandaru, H. (2021). *Pemodelan basis data relasional*. Yogyakarta: Deepublish.
- [13] Wiranata, I. G. (2020). *Perancangan basis data*. Jakarta: Andi