

# Penerapan Algoritma K-Means pada Sistem Pencarian Gambar E-Repository Prodi Ti UNIMA

<sup>1</sup>Jibrael Muntuuntu, <sup>2</sup>Glenn D. P. Maramis, <sup>3</sup>Quido C. Kainde  
<sup>1,2,3</sup>Universitas Negeri Manado,  
Tondano, Indonesia

<sup>1</sup>[muntuunturael@gmail.com](mailto:muntuunturael@gmail.com), <sup>2</sup>[gmaramis@unima.ac.id](mailto:gmaramis@unima.ac.id), <sup>3</sup>[quidokainde@unima.ac.id](mailto:quidokainde@unima.ac.id)

\*Penulis Korespondensi

Diajukan : 20/11/2025

Diterima : 09/01/2026

Dipublikasi : 13/01/2026

## ABSTRAK

Pengelolaan karya ilmiah di Program Studi Teknik Informatika Universitas Negeri Manado masih didominasi arsip fisik, menyebabkan aset visual penting seperti diagram dan *flowchart* sulit ditemukan. Penelitian ini bertujuan membangun sistem repositori digital cerdas yang mampu mengekstrak teks dari gambar menggunakan *Optical Character Recognition* (OCR) dan mengelompokkan dokumen menggunakan algoritma K-Means. Populasi dalam penelitian ini adalah seluruh dokumen skripsi yang tersimpan di repositori program studi. Teknik pengambilan sampel menggunakan *purposive sampling* untuk mengumpulkan dataset gambar diagram dan *flowchart* dari dokumen tahun 2020-2024 sebagai data uji. Metode penelitian menggunakan *Agile Software Development* dengan *framework* Laravel 12, Tesseract OCR, dan *library* PHP-ML. Hasil penelitian menunjukkan bahwa fitur pencarian berbasis gambar berhasil diimplementasikan, di mana sistem mampu mengekstrak teks dari gambar unggahan pengguna sebagai kata kunci pencarian. Selain itu, algoritma K-Means berhasil dijalankan sebagai proses *backend* untuk mengelompokkan dokumen ke dalam 10 kluster berdasarkan kemiripan teks, yang dibuktikan dengan terisinya kolom *cluster\_id* pada basis data. Sistem ini berhasil mentransformasi repositori fisik menjadi sistem digital yang mampu mengindeks konten visual yang sebelumnya tidak terakses.

**Kata Kunci:** *Clustering*, K-Means, OCR, Pencarian Gambar, Repositori Elektronik.

## I. PENDAHULUAN

Perkembangan teknologi informasi mendorong institusi pendidikan beralih ke sistem digital untuk mempermudah akses informasi bagi pengguna (Supriyanto, n.d.). Penerapan sistem berbasis web semacam ini telah terbukti efektif dalam meningkatkan layanan perpustakaan di berbagai institusi (Ika Rahmawati Suci, Noviam Adi Prasetyo, Gita Fadila Fitriana, 2021). Meskipun demikian, di Universitas Negeri Manado (UNIMA), pengelolaan karya ilmiah seperti skripsi sebagian besar masih mengandalkan arsip dalam bentuk fisik. Kondisi ini menjadikan repositori hanya berfungsi sebagai gudang dokumen, bukan basis data yang mudah diakses. Masalah fundamental yang dihadapi adalah sulitnya menemukan konten visual seperti diagram sistem dan *flowchart* algoritma yang "terkubur" di dalam tumpukan arsip fisik. Metode pencarian manual saat ini tidak memungkinkan penemuan informasi spesifik yang terdapat di dalam gambar tersebut.

Kesenjangan penelitian (*research gap*) terlihat dari studi-studi sebelumnya yang fokus pada digitalisasi teks namun belum menyentuh aspek ekstraksi konten visual. Diperlukan solusi yang tidak hanya menyimpan dokumen, tetapi juga mampu "membaca" konten gambar di dalamnya. Tujuan penelitian ini adalah merancang sistem repositori yang menerapkan teknologi *Optical Character Recognition* (OCR) untuk ekstraksi teks dari gambar dan algoritma K-Means untuk pengelompokan dokumen, guna meningkatkan efisiensi pencarian referensi visual.

## II. STUDI LITERATUR

### Penelitian Terdahulu

Untuk mendukung penelitian ini, dilakukan penelaahan terhadap beberapa studi terdahulu yang relevan. Penelitian oleh (Omega et al., n.d.) mengembangkan sistem repositori skripsi berbasis web di Teknik Informatika UNIMA menggunakan metode Agile, yang berhasil mendigitalisasi manajemen judul namun belum memiliki fitur pencarian konten gambar. Selanjutnya, (Hasanah & Purnomo, 2022) menerapkan algoritma K-Means untuk mengelompokkan buku perpustakaan berdasarkan frekuensi peminjaman, membuktikan efektivitas algoritma ini dalam manajemen koleksi.

Senada dengan itu, (Yulia & Silalahi, 2021) juga menggunakan K-Means untuk klusterisasi buku, yang membantu dalam pengambilan keputusan pengadaan. Selain itu, penelitian oleh (Alvionita Manikome & David Paulus Maramis, n.d.) juga telah menerapkan K-Means untuk pengarsipan surat, sementara (Subekti & Pratama, 2024) menggunakannya untuk memprediksi kelulusan mahasiswa. Penelitian ini melengkapi studi terdahulu dengan menggabungkan K-Means dan teknologi OCR Tesseract.

**Algoritma K-Means**, K-Means adalah metode *clustering* non-hierarkis yang mempartisi data ke dalam  $k$  kelompok. Algoritma ini bekerja dengan meminimalkan varians antara data dalam satu kluster dan memaksimalkan varians antar kluster (Afifah & Nurdianto, 2023; Nasari et al., 2022).

**Optical Character Recognition (OCR)**, OCR adalah teknologi yang mengonversi citra huruf menjadi karakter teks yang dapat dikenali mesin. Dalam penelitian ini, *engine* Tesseract OCR digunakan untuk mengubah piksel pada gambar diagram menjadi teks yang dapat diindeks.

## III. METODE

Populasi dan Sampel Populasi penelitian ini mencakup seluruh dokumen skripsi, proposal, dan jurnal di Program Studi Teknik Informatika UNIMA. Sampel data uji berupa kumpulan file gambar (diagram, flowchart) yang diekstraksi dari dokumen skripsi periode 2020-2024 untuk pengujian akurasi OCR dan clustering.

Teknik Pengumpulan dan Analisis Data Penelitian ini menggunakan metode Agile Software Development (Abrahamsson et al., 2002). Metode ini dipilih karena pendekatannya yang iteratif dan inkremental, yang berfokus pada fleksibilitas, kolaborasi tim, dan kemampuan beradaptasi terhadap perubahan kebutuhan selama proses pengembangan berlangsung (Hariyanto et al., 2024). Pendekatan Agile menekankan pada pengiriman perangkat lunak yang cepat dan responsif terhadap umpan balik pengguna (Dzaky & Kurniawan, 2023; Nakash, 2024). Tahapan pengembangan sistem ini dapat dilihat pada Gambar 1.

Gambar 1. Metode *Agile Software Development*



(Omega et al., n.d.)

Pengumpulan data dilakukan melalui observasi proses pencarian manual, wawancara dengan mahasiswa dan dosen, serta studi pustaka. Sistem dibangun menggunakan *framework* Laravel 12 dan basis data MySQL yang berjalan pada server berbasis PHP (Wulan & Nudin, n.d.) Analisis data tekstual untuk K-Means dilakukan melalui tahap *preprocessing* (pembersihan teks), pembobotan TF-IDF, dan pengelompokan menggunakan *library* PHP-ML. Pengujian sistem dilakukan menggunakan metode *Black Box Testing* untuk memvalidasi fungsionalitas input-output.

Perancangan alur sistem divisualisasikan menggunakan diagram *Use Case* (Fernando et al., 2024) sebagaimana ditunjukkan pada **Gambar 2**.

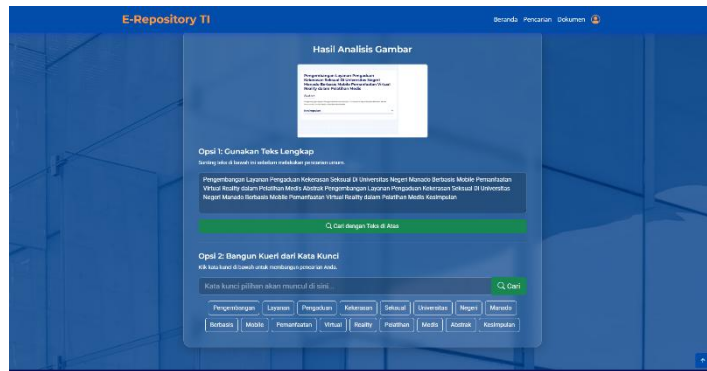


Gambar 2. Use Case Diagram

**IV. HASIL DAN PEMBAHASAN**

Penelitian ini menghasilkan sistem E-Repository dengan dua fitur kecerdasan utama, pencarian berbasis gambar dan pengelompokan dokumen otomatis.

**Implementasi Pencarian Berbasis Gambar (OCR)**, Sistem menyediakan antarmuka bagi pengguna untuk mengunggah gambar. *Engine* Tesseract OCR memproses gambar tersebut dan mengekstrak teks di dalamnya menjadi kata kunci, sebagaimana ditampilkan pada **Gambar 3**. Berdasarkan pengujian *Black Box*, fitur ini berfungsi baik pada gambar dengan resolusi jelas, memungkinkan pengguna menemukan dokumen sumber dari sebuah potongan gambar diagram.



Gambar 3. Hasil Analisa OCR

**Implementasi Clustering K-Means**, Proses K-Means dijalankan sebagai perintah latar belakang (*background job*). Tabel 1 menunjukkan hasil pengujian eksekusi algoritma ini.

Tabel 1. Hasil Pengujian K-Means

No	Skenario Uji	Input	Output yang diharapkan	Output Sistem	Status
1.	Eksekusi perintah K-Measn (data tersedia)	- Menjalankan perintah php artisan app:process-kmeans-clustering pada terminal server.	- perintah berjalan tanpa error - konsol menampilkan log proses: "memulai....", "mengumpulkan data...", "menjalankan K-Means...", "menyimpan hasil...".	Perintah selesai dengan log proses serta pesan sukses sesuai harapan.	Berhasil

No	Skenario Uji	Input	Output yang diharapkan	Output Sistem	Status
			-proses diakhiri dengan pesan “Prosesn clustering K-Means berhasil diselesaikan!”.		
2.	Verifikasi hasil di database	- (setelah skenario 1 berhasil) - memeriksa tabel skripsi, proposals, dan jurnals, di database.	- kolom cluster_id pada semua baris data telah terisi dengan nilai numerik atau ID cluster.	Kolon cluster_id terisi dengan angka integer	Berhasil
3.	Eksekusi perintah (data kosong)	- Mengosongkan tabel skripsi, proposals, dan jurnals - menjalankan perintah php artisan app:process-kmeans-clustering	- Perintah berjalan dan berhenti secara graceful. - konsol menampilkan pesan “Tidak ada dokumen untuk diproses.”	- Perintah berhenti dengan pesan “Tidak ada dokumen untuk di proses”.	Berhasil

Sumber tabel: Data Olahan Peneliti (2025)

Seperti terlihat pada Tabel 1, sistem berhasil mengisi kolom cluster\_id pada basis data. Hal ini membuktikan bahwa dokumen-dokumen telah berhasil dikelompokkan ke dalam 10 klaster topik yang berbeda, yang dibuktikan dengan data pada **Gambar 4**.

cluster_id	nama_penulis	tahun	judul	abstrak	kata_kunci
0	Enjelina	2025	ANALISIS CLUSTERING TINGKAT KEPARAHAN PENYAKIT PA...	Rumah sakit Helirmana Leimbelan merupakan salah sa...	K-Means, Penyakit, Clustering
5	Rafter Johannes Karouw	2025	APLIKASI PENGELOLAAN DATA PRODUKSI DAN KONSUMSI BE...	Bahan pangan merupakan kebutuhan pokok masyarakat ...	Aplikasi, Waterfall
9	Faldy A. Gosal	2025	APLIKASI PREDIKSI NILAI TUKAR RUPIAH TERHADAP DOLL...	Fluktuasi nilai tukar Rupiah terhadap Dolar Amerik...	Prediksi nilai tukar, Autoregressive Integrated Mo...

Gambar 4. Hasil Clustering K-Means

Keberhasilan integrasi OCR menjawab masalah kesulitan pencarian aset visual yang dipaparkan pada bagian pendahuluan. Berbeda dengan sistem repositori konvensional, sistem ini memungkinkan penelusuran konten *di dalam* gambar. Penerapan K-Means juga memberikan nilai tambah berupa pengorganisasian dokumen secara otomatis, yang menurut saran pengguna pada tahap umpan balik, dapat dikembangkan lebih lanjut menjadi fitur "Rekomendasi Dokumen Serupa".

#### IV. KESIMPULAN

Penelitian ini berhasil merancang dan membangun sistem E-Repository cerdas menggunakan metode *Agile*. Sistem sukses mengimplementasikan teknologi OCR yang memungkinkan pencarian dokumen berbasis konten gambar, serta menerapkan algoritma K-Means yang terbukti mampu mengelompokkan dokumen skripsi ke dalam kluster-kluster tematik secara otomatis. Sistem ini secara efektif mengatasi keterbatasan aksesibilitas arsip fisik dan meningkatkan efisiensi penemuan referensi visual bagi sivitas akademika.

#### V. REFERENSI

- Abrahamsson, A. P., Salo, O., & Ronkainen, J. (2002). *Agile Software Development Methods : Review and Analysis*.
- Afifah, I. A. N., & Nurdiyanto, H. (2023). Data Mining Clustering Dalam Pengelompokan Buku Perpustakaan Menggunakan Algoritma K-Means. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 8(3), 802–814. <https://doi.org/10.29100/jipi.v8i3.3891>
- Alvionita Manikome, H., & David Paulus Maramis, G. (n.d.). *Arsip Surat Masuk Dan Keluar Pada Kejaksaan Tinggi Sulawesi Utara Dengan Algoritma K-Means Berbasis Web*.
- Dzaky, F. A., & Kurniawan, D. (2023). *Implementasi Metode Agile Framework Scrum dalam Pengembangan Sistem Informasi Manajemen Aset Terpadu Universitas Diponegoro Modul Inventarisasi* (Vol. 14, Issue 1).
- Fernando, N., Tua, S., & Effendy, I. (2024). *Aplikasi Helpdesk Ticketing berbasis Website pada PT Pertamina EP Limau Field Zona 4 dengan Metode Agile*. 10(2), 616–631.
- Hariyanto, D., Rivaldy, M., Ricki, Y., Indonesia, U. S., Bina, U., & Informatika, S. (2024). *Implementasi SIMDUMAS Metode Agile Software Development Extreme Programming Pada Kecamatan Bogor Barat Kota Bogor*. 5(2), 31–35.
- Hasanah, N. N., & Purnomo, A. S. (2022). Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering (Studi Kasus : Perpustakaan Politeknik LPP Yogyakarta). *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 4(2), 300–311. <https://doi.org/10.47233/jteksis.v4i2.499>
- Ika Rahmawati Suci, Noviam Adi Prasetyo, Gita Fadila Fitriana. (2021). *Buku Tamu Perpustakaan Berbasis Website Dengan Metode Agile (Perpustakaan Badan Pusat Statistik Kabupaten Brebes)*. 7(2), 70–76.
- Nakash, M. (2024). *Agile Software Development : The Experience of Working in Sprints A GILE S OFTWARE D EVELOPMENT : July*. <https://doi.org/10.28945/5252>
- Nasari, F., Sianturi, C. J. M., Azis, W. S., dan Dedy Atmajaya, Ulfah, M., Irtwaty, A. S., Triyansyah, D., Fitriana, D., Afifah, I. A. N., Nurdiyanto, H., Suraya, S., Sholeh, M., Andayati, D., Priyatman, H., Sajid, F., Haldivany, D., JASMINE, K., Kodinariya, T. M., Makwana, P. R., ... Rachmayani, A. N. (2022). No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title. *Libraria Jurnal Ilmu Perpustakaan Dan Informasi*, 4(2), 1–12. <https://doi.org/10.36080/skanika.v6i1.2982>

- 
- Omega, C., Pengembangan, R. :, Repositori, S., Rimbing, C., Rorimpandey, G., & Rantung, V. (n.d.). PENGEMBANGAN SISTEM REPOSITORI SKRIPSI DI TEKNIK INFORMATIKA UNIVERSITAS NEGERI MANADO BERBASIS WEB. In *JOINTER: JOURNAL OF INFORMATICS ENGINEERING* (Vol. 04, Issue 02).
- Subekti, P., & Pratama, A. (2024). *Analisis dan Perancangan Sistem Informasi Perpustakaan Digital Berbasis Web*. 2(2), 70–79.
- Supriyanto, W. (n.d.). *PENGEMBANGAN SISTEM LAYANAN PERPUSTAKAAN DIGITAL*.
- Wulan, D. R., & Nudin, S. R. (n.d.). *Rancang Bangun Sistem Informasi Persuratan berbasis Web di Fakultas Vokasi Universitas Negeri Surabaya*. 1–11.
- Yulia, & Silalahi, M. (2021). Penerapan Data Mining Clustering Dalam Mengelompokan Buku Dengan Metode K-Means. *Indonesian Journal of Computer Science*, 10(1). <https://doi.org/10.33022/ijcs.v10i1.3008>