
IMPLEMENTASI ALGORITMA K-MEANS UNTUK CLUSTERING DOKUMEN HUKUM DALAM SISTEM PENCARIAN DOKUMEN DI KANTOR HUKUM

Wan Khaira Afwika¹, Muhammad Yasin S²

Universitas Asahan, Asahan

e-mail: wankhaira401@gmail.com¹, muhammadyasins@gmail.com²

Abstract: *Law firms are institutions that rely heavily on legal documents as sources of information and as a basis for decision-making. Over time, the number of stored legal documents, such as contracts, powers of attorney, court decisions, and laws and regulations, increases in number and variety. This often makes it difficult to quickly and accurately find relevant documents, especially if document management is not systematically structured. Manual document searches are not only time-consuming but also prone to human error, such as errors in categorization or indexing. The data used in this study are legal documents in text form, such as court decisions, contracts, laws and regulations, and other similar documents in digital format (.txt, .docx, or .pdf files that have been converted to text). The purpose of this study is to implement the K-Means algorithm to cluster legal documents based on their content. More relevant and faster search results with clustering, such as cluster 0 documents on criminal law, corruption, and restorative justice; cluster 1 documents on personal data protection, electronic information technology (ITE), and online gender-based violence; and cluster 2 documents on civil and administrative law, such as employment agreements, consumer protection, and customary land. The system, integrated with clustering results, shows improved performance compared to conventional search systems that rely solely on keyword searches.*

Keywords: *Implementation, K-Means Algorithm, Clustering, Legal Documents. Web*

Abstrak: Kantor hukum merupakan institusi yang sangat bergantung pada dokumen hukum sebagai sumber informasi dan dasar pengambilan keputusan. Seiring berjalannya waktu, jumlah dokumen hukum yang tersimpan seperti kontrak, surat kuasa, putusan pengadilan, dan peraturan perundang-undangan semakin banyak dan beragam. Kondisi ini sering kali menimbulkan kesulitan dalam pencarian dokumen yang relevan secara cepat dan tepat, terutama jika pengelolaan dokumen tersebut belum terstruktur secara sistematis. Pencarian dokumen secara manual tidak hanya memakan waktu, tetapi juga rentan terhadap kesalahan manusia, seperti kekeliruan dalam pengkategorian atau pengindeksan. Data yang digunakan dalam penelitian ini ialah dokumen hukum berbentuk teks, seperti putusan pengadilan, kontrak, peraturan perundang-undangan, dan dokumen sejenis lainnya dalam format digital (.txt, .docx, atau .pdf yang sudah diubah menjadi teks). Tujuan dalam penelitian ini mengimplementasikan algoritma K-Means untuk melakukan clustering terhadap dokumen hukum berdasarkan isi dokumen. Hasil pencarian yang lebih relevan dan cepat dengan pembagian cluster seperti cluster 0 dokumen bertema hukum pidana, korupsi, dan keadilan restoratif, cluster 1 Dokumen bertema perlindungan data pribadi, ITE, dan kekerasan berbasis gender online dan cluster 2 Dokumen bertema hukum perdata dan administrasi, seperti perjanjian kerja, konsumen, dan tanah adat. Sistem yang telah diintegrasikan dengan hasil clustering menunjukkan peningkatan performa dibandingkan sistem pencarian konvensional yang hanya mengandalkan pencarian kata kunci.

Kata Kunci: Implementasi, Algoritma K-Means, Clustering, Dokumen Hukum. Web

PENDAHULUAN

Perkembangan teknologi informasi yang pesat dalam dua dekade terakhir telah menghasilkan data explosion dalam berbagai bidang, seperti bisnis, kesehatan, pendidikan, keuangan, dan hukum. Volume data yang terus meningkat menuntut adanya teknologi yang mampu mengolah dan mengekstraksi informasi penting dari kumpulan data yang besar dan kompleks. Dalam konteks ini, data mining telah menjadi salah satu bidang penting yang berkembang pesat dan banyak diterapkan dalam pengambilan keputusan berbasis data.

Data mining merupakan suatu proses yang kompleks dengan menggunakan statistik, matematika, kecerdasan buatan, dan *machine learning* guna mengekstraksi informasi yang berharga dari *database* berskala besar. Informasi yang terhimpun diolah dalam *database* untuk mendukung pengambilan keputusan (Riky Maoulana et al., 2024). Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakut dari berbagai *database* besar (Kusnadi et al., 2021).

Kantor hukum Zulkifli yang beralamatkan Jl. Sisingamangaraja Nomor 270 Kelurahan Tegal Sari Kecamatan Kota Kisaran Barat Kabupaten Asahan merupakan institusi yang sangat bergantung pada dokumen hukum sebagai sumber informasi dan dasar pengambilan keputusan. Seiring berjalannya waktu, jumlah dokumen hukum yang tersimpan seperti kontrak, surat kuasa, putusan pengadilan dan peraturan perundang-undangan semakin banyak dan beragam. Permasalahan menimbulkan kesulitan dalam pencarian dokumen yang relevan secara cepat dan tepat, terutama jika pengelolaan dokumen tersebut belum terstruktur secara sistematis. Pencarian dokumen secara manual tidak hanya memakan waktu, tetapi juga rentan

terhadap kesalahan manusia, seperti kekeliruan dalam pengkategorian atau pengindeksan.

Salah satu solusi yang dapat diterapkan adalah dengan menggunakan teknik data mining, khususnya metode clustering untuk mengelompokkan dokumen. Metode K-Means merupakan algoritma clustering yang banyak digunakan karena kemampuannya dalam membagi data ke dalam sejumlah kelompok (*cluster*) berdasarkan kemiripan fitur. Dengan mengimplementasikan algoritma K-Means pada kumpulan dokumen hukum, maka sistem dapat mengelompokkan dokumen ke dalam beberapa kategori seperti pidana, perdata, ketenagakerjaan, dan sebagainya secara otomatis.

Clustering atau analisis kelompok merupakan proses mengelompokkan data (objek) berdasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan diantaranya (Muhammad Norshahlan et al., 2023). *Clustering* atau klasterisasi merupakan suatu teknik atau metode untuk mengelompokkan data. *Clustering* adalah salah satu alat yang penting dalam pengolahan data statistik untuk melakukan analisis data (Nur Afidah, 2023).

K-Means adalah salah satu metode analisis data yang digunakan untuk mengelompokkan data menjadi beberapa kelompok yang saling berhubungan berdasarkan kesamaan atribut (Maoulana et al., 2024). Algoritma *K-Means* merupakan algoritma pengelompokan *iterative* yang melakukan partisi set data kedalam sejumlah *cluster* yang sudah ditetapkan di awal. Metode ini meminimalkan berbagai bentuk perbedaan yang terjadi antar data didalam satu *cluster* dan juga memaksimalkan perbedaan yang terjadi dengan *cluster* yang lain (Abdul Arif et al., 2022).

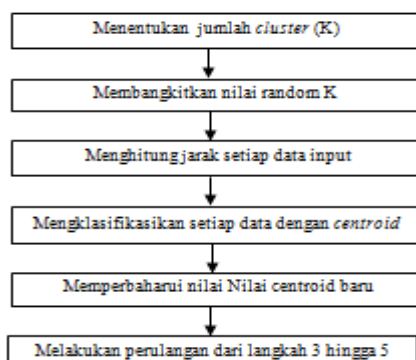
Penelitian terdahulu yang pernah dilakukan oleh (Fahrul Fauji et al., 2025) dengan judul "Penerapan Data Mining Menggunakan K-Means Clustering Dalam Mengelompokkan Tingkat

Kesulitan Mata Pelajaran” Hasil penelitian setelah 163 data ditransformasi menjadi 30 data didapatkan 3 cluster optimal melalui metode Elbow, yaitu Cluster_1 dengan kategori pelajaran sulit berisi 8 mata pelajaran. Cluster_2 dengan kategori pelajaran tingkat kesulitan sedang berisi 10 matapelajaran. Cluster_3 dengan kategori pelajaran mudah berisi 12 mata pelajaran. Hasil pengelompokan ini dapat digunakan para guru di SMA Kertanegara Malang untuk memberikan pendampingan lebih kepada siswa, terutama pada mata pelajaran yang dikategorikan sulit.

Penelitian terdahulu yang pernah dilakukan oleh (Ramadana et al., 2025) dengan judul” Implementasi Algoritma K-Means Clustering terhadap Tingkat Kepuasan Peserta LKP Multi Talenta Komputer Purwokerto” berdasarkan pengujian jumlah data yang masuk dalam kategori Kurang adalah 3 data dengan nilai rata-rata 4,9. Sedangkan data yang masuk pada kategori Cukup adalah 28 data dengan rata-rata 3,9. Sedangkan jumlah data yang masuk dalam kategori baik berjumlah 24 data dengan nilai rata-rata 4,3.

METODE

Kerangka kerja penelitian ini penjabaran dari langkah-langkah yang akan dilakukan dalam penyelesaian masalah dengan menggunakan metode K-Means. Adapun kerangka kerja penelitian ini dapat dilihat pada Gambar 1 berikut:



Gambar 1 Kerangka Kerja penelitian

Penelitian ini merupakan penelitian kualitatif dengan rancangan studi kasus, alasan penelitian kualitatif digunakan peneliti karena dalam penelitian kualitatif cenderung menggunakan analisis dalam menyelesaikan suatu masalah. Dalam pengumpulan data ini, metode yang penulis gunakan berupa metode observasi yang nantinya digunakan untuk memperoleh data-data yang dibutuhkan dalam pembuatan aplikasi. Metode observasi ini berupa pengumpulan data dari sumber-sumber yang berkaitan dengan perancangan aplikasi, antara lain seperti clustering dokumen hukum dalam sistem pencarian dokumen di kantor hukum, Metode penelitian merupakan suatu langkah yang paling menentukan dari suatu penelitian, karena analisa data berfungsi untuk menyimpulkan hasil penelitian.

HASIL DAN PEMBAHASAN

Hasil Penelitian

Analisa Data

Data yang digunakan dalam penelitian ini terdiri atas sepuluh dokumen hukum yang diunggah ke sistem “Sistem Dokumen Hukum”. Setiap dokumen berisi teks yang bersumber dari kajian atau analisis terhadap berbagai undang-undang dan peraturan di Indonesia, antara lain sebagai berikut:

Tabel 1 Jenis Dokumen

No	Dokumen
1	Implementasi Restorative Justice
2	Penyalahgunaan Wewenang
3	Asas Legalitas
4	Perlindungan Data Pribadi
5	Undang-Undang ITE
6	Kekerasan Gender Online
7	Perjanjian Kerja Cipta Kerja
8	Perlindungan Konsumen
9	Sengketa Tanah Adat
10	Surat Kuasa
11	Cerai Gugat

Atribut data dalam tabel dokumen terdiri dari:

1. id : identitas dokumen
2. judul : judul dokumen hukum
3. isi : isi atau ringkasan dokumen hukum
4. tanggal_upload : waktu dokumen dimasukkan ke sistem
5. cluster_id : hasil pengelompokan dokumen berdasarkan kemiripan isi
6. file_path : lokasi file (PDF/DOC) yang diunggah

Pada tahap awal, nilai cluster_id dan file_path masih kosong. Nilai cluster_id akan diisi setelah proses clustering selesai dilakukan.

Tahapan Analisis Data

Tahapan analisis dokumen hukum dilakukan secara otomatis melalui skrip clustering.php dengan urutan proses sebagai berikut:

Preprocessing (Praproses Teks)

1. Semua huruf dikonversi menjadi huruf kecil (*lowercase*).
2. Simbol dan tanda baca dihapus menggunakan ekspresi reguler $[/^a-zs]/$.
3. Dokumen dipisahkan menjadi token (kata-kata).
4. Stopwords umum dalam bahasa Indonesia (seperti “dan”, “yang”, “pada”, “untuk”, dll) dihapus agar hanya kata bermakna yang tersisa.

Contoh Praproses:

Dokumen:

“Penelitian ini membahas penerapan Undang-Undang Nomor 19 Tahun 2016 tentang Informasi dan Transaksi Elektronik (ITE).”

Setelah praproses menjadi:

['penelitian','membahas','penerapan','undang-undang','nomor','informasi','transaksi','elektronik','ite']

Pembentukan Vocabulary (Kumpulan Kata Unik)

Dari seluruh dokumen yang telah diproses, dibuat daftar kata unik

(*vocabulary*) yang digunakan untuk membentuk representasi vektor.

Perhitungan TF (Term Frequency) TF menunjukkan jumlah kemunculan kata dalam satu dokumen.

$$TF(t, d) = \frac{\text{Jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{Jumlah total kata dalam dokumen } d}$$

Contoh Perhitungan Manual:

Misal dalam dokumen 1 terdapat 100 kata, dan kata “undangundang” muncul 5 kali, maka:

$$TF(\text{undangundang}, d_1) = 5/100 = 0.05$$

Perhitungan IDF (*Inverse Document Frequency*) IDF digunakan untuk menurunkan bobot kata yang muncul terlalu sering di banyak dokumen.

$$IDF(t) = \log \frac{N + 1}{df_t + 1} + 1$$

di mana:

N = jumlah total dokumen

df_t = jumlah dokumen yang mengandung kata t

Contoh:

Kata “undangundang” muncul di 7 dari 10 dokumen, maka:

$$IDF(\text{undangundang}) = \log \frac{10 + 1}{7 + 1} + 1 = \log(1.375) + 1 = 1.137$$

Perhitungan TF-IDF Setiap kata dalam dokumen diberi bobot:

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

Contoh:

$TF(\text{undangundang}, d_1) = 0.05$ dan $IDF(\text{undangundang}) = 1.137$

Maka:

$$TFIDF(\text{undangundang}, d_1) = 0.05 \times 1.137 = 0.05685$$

Perhitungan Jarak Euclidean (Kemiripan Dokumen)

Setiap dokumen direpresentasikan sebagai vektor TF-IDF. Jarak antar dua dokumen dihitung dengan rumus:

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

Contoh:

Dokumen 1 $\rightarrow [0.05, 0.02, 0.01]$

Dokumen 2 $\rightarrow [0.04, 0.01, 0.03]$

$$d = \sqrt{(0.05 - 0.04)^2 + (0.02 - 0.01)^2 + (0.01 - 0.03)^2}$$

$$= \sqrt{0.0001 + 0.0001 + 0.0004}$$

$$= \sqrt{0.0006}$$

$$= 0.0245$$

Semakin kecil jarak, semakin mirip kedua dokumen.

Penerapan Algoritma K-Means

1. Jumlah cluster (k) ditentukan sebanyak 3.
2. Sistem memilih 3 dokumen awal secara acak sebagai *centroid* (pusat cluster).
3. Setiap dokumen kemudian dihitung jaraknya ke masing-masing *centroid*, dan dimasukkan ke cluster terdekat.
4. Setelah itu, *centroid* baru dihitung sebagai rata-rata vektor TF-IDF semua anggota cluster.
5. Proses diulang hingga perubahan nilai *centroid* sangat kecil (*konvergen*).

Perhitungan K-Means

Misalkan tiga dokumen pertama setelah diubah menjadi vektor TF-IDF seperti berikut (dalam 3 dimensi untuk penyederhanaan):

Tabel 2 Cluster

Dokumen	Kata	Kata	Kata
---------	------	------	------

Tabel 3 Hasil Clustering

Cluster	Karakteristik Umum	Dokumen
Cluster 0	Dokumen bertema hukum pidana, korupsi, dan keadilan restorative	“Implementasi Restorative Justice”, “Penyalahgunaan Wewenang”, “Asas Legalitas”
Cluster 1	Dokumen bertema perlindungan data pribadi, ITE, dan kekerasan berbasis gender online	“Perlindungan Data Pribadi”, “Undang-Undang ITE”, “Kekerasan Gender Online”
Cluster 2	Dokumen bertema hukum perdata dan administrasi, seperti perjanjian kerja, konsumen, dan tanah adat	“Perjanjian Kerja Cipta Kerja”, “Perlindungan Konsumen”, “Sengketa Tanah Adat”

Analisis Hasil

Hasil pengelompokan menunjukkan bahwa sistem berhasil memisahkan dokumen berdasarkan kemiripan makna dan konteks hukum. Kata-kata yang dominan pada masing-masing cluster menentukan kedekatan dokumen:

	A	B	C
D1	0.10	0.05	0.00
D2	0.11	0.04	0.01
D3	0.02	0.08	0.10

Pemilihan Awal Centroid:

$$C_1 = D1, C_2 = D2, C_3 = D3$$

Hitung jarak D1 ke semua centroid:

1. $d(D1, C_1) = 0$
2. $d(D1, C_2) = \sqrt{((0.10-0.11)^2 + (0.05-0.04)^2 + (0-0.01)^2)} = 0.017$
3. $d(D1, C_3) = \sqrt{((0.10-0.02)^2 + (0.05-0.08)^2 + (0-0.10)^2)} = 0.125$

D1 masuk ke cluster 1 (jarak paling kecil). Proses yang sama dilakukan untuk dokumen lain.

Hitung Centroid Baru

Setelah semua dokumen masuk ke cluster tertentu, *centroid* baru dihitung dengan mengambil rata-rata dari semua vektor di dalam cluster.

Ulangi hingga konvergen.

Hasil Clustering

Berdasarkan hasil eksekusi sistem pada 11 dokumen hukum diperoleh pembagian menjadi 3 cluster sebagai berikut:

1. Cluster 0 banyak memuat kata “pidana”, “keadilan”, “korupsi”, “wewenang”.
2. Cluster 1 mengandung kata “data”, “pribadi”, “informasi”, “elektronik”.
3. Cluster 2 mengandung kata “kerja”, “tanah”, “konsumen”, “perjanjian”.

Visualisasi distribusi hasil clustering juga dapat dilihat melalui diagram batang yang ditampilkan menggunakan Chart.js pada halaman hasil clustering.

Pembahasan

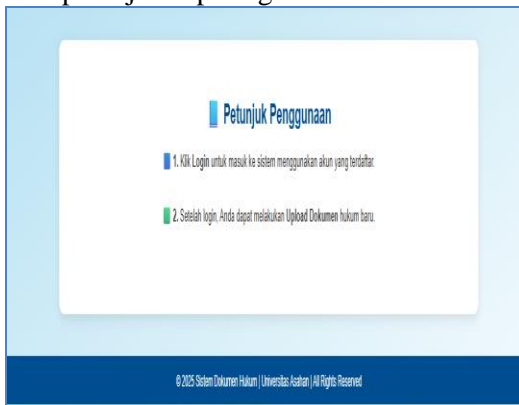
Dalam menjalankan sistem yang penulis buat, berikut langkah-langkah yang harus dilaksanakan :

1. Membuka aplikasi yang sudah dipersiapkan. maka akan tampil halaman menu awal, berikut ini tampilan menu awal seperti gambar 2.



Gambar 2 Halaman Tampilan Menu Awal

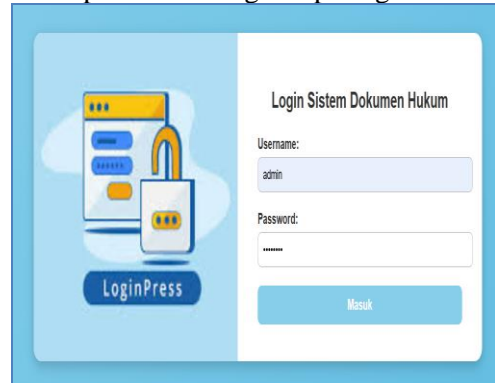
2. Klik menu petunjuk yang terdapat pada menu awal, menu ini digunakan untuk memberikan petunjuk tentang penggunaan aplikasi yang peneliti buat, berikut ini tampilan menu petunjuk seperti gambar 3.



Gambar 3 Halaman Tampilan Menu Petunjuk

3. Klik menu login yang terdapat pada menu awal, menu ini digunakan untuk proses login ke menu utama dengan menginputkan username dan password

pengguna atau admin, berikut ini tampilan menu login seperti gambar 4.



Gambar 4 Halaman Tampilan Login

4. Setelah proses login selesai. maka akan tampil halaman menu utama admin seperti gambar 5.



Gambar 5 Halaman Tampilan Utama Admin

5. Klik menu upload dokumen yang terdapat pada menu utama admin kemudian upload dokumen untuk menambah data maupun melakukan aksi edit dan hapus data, menu upload dokumen seperti terlihat pada gambar 6.

ID	Judul	Isi Singkat	Tanggal Upload	File	Aksi
10	Perencanaan Ases Legalitas dalam Hukum Pidana Indonesia	Artikel ini menganalisis makna ases legalitas menurut...	2025-10-10 00:00:00	Lihat Hapus	Edit Hapus
9	Analisis Yuridis Sengketa Tanah Adat di Indonesia	Tulisan ini menjelaskan konsep tanah adat menurut...	2025-10-09 00:00:00	Lihat Hapus	Edit Hapus
8	Tinjauan Hukum Pidana terhadap Kekerasan Berbasis Gender Online	Makalah ini membahas bentuk kekerasan berbasis gen...	2025-10-08 00:00:00	Lihat Hapus	Edit Hapus
7	Pertanggungjawaban Pidana Korporasi dalam Kasus Lingkungan Hidup	Penelitian ini mengaji penerapan pasal 116 dan 11...	2025-10-07 00:00:00	Lihat Hapus	Edit Hapus
6	Perlindungan Hukum Terhadap Konsumen dalam Transaksi E-Commerce	Artikel ini menganalisis mekanisme perlindungan ko...	2025-10-06 00:00:00	Lihat Hapus	Edit Hapus
5	Analisis Hukum Perjanjian Kerja Berdasarkan Undang-Undang Cipta Kerja	Kajian ini menyoroti perubahan substansi hukum ke...	2025-10-05 00:00:00	Lihat Hapus	Edit Hapus
4	Tinjauan Yuridis terhadap Penyahgunaan Wewenang dalam Jabatan Publik	Tulisan ini mengulas unsur-unsur tindak pidana pen...	2025-10-04 00:00:00	Lihat Hapus	Edit Hapus
3	Implementasi Restorasi Ekosistem dalam Penegakan Hukum Pidana di Indonesia	Makalah ini membahas konsep keadilan restoratif da...	2025-10-03 00:00:00	Lihat Hapus	Edit Hapus
2	Kajian Hukum Perlindungan Data Pribadi Berdasarkan UU No. 27 Tahun 2022	Dokumen ini menjelaskan substansi Undang-Undang Pr...	2025-10-02 00:00:00	Lihat Hapus	Edit Hapus
1	Analisis Penerapan Undang-Undang ITE dalam Kasus Percekcupan Nama Bank di Media Sosial	Penelitian ini membahas penerapan Undang-Undang I...	2025-10-01 00:00:00	Lihat Hapus	Edit Hapus

Gambar 6 Halaman Menu Upload Dokumen

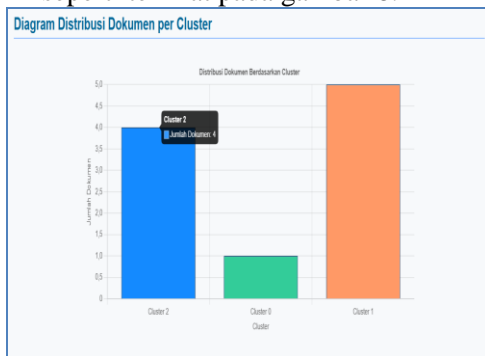
6. Klik menu Lakukan clustering yang terdapat pada menu utama admin sistem akan secara otomatis

melakukan clustering jenis dokumen hukum, menu proses clustering seperti terlihat pada gambar 7.

No	Judul Dokumen	Cuplikan Isi	Tanggal Upload
1	Analisis Penerapan Undang-Undang ITE dalam Kasus Pencemaran Nama Baik di Media Sosial	Penelitian ini membahas penerapan Undang-Undang Nomor 19 Tahun 2016 tentang Informasi dan Transaksi Elektronik (ITE) terhadap kasus pencemaran nama baik di media sosial. Analisis dilakukan terhadap pasal 27 ayat (3) dan pasal 45 ayat (3), serta perba...	2025-10-01 00:00:00
2	Implementasi Restorative Justice dalam Penegakan Hukum Pidana di Indonesia	Makalah ini membahas konsep keadilan restoratif dalam sistem hukum pidana Indonesia, terutama penerapan Peraturan Kapolri Nomor 8 Tahun 2021. Pembahasan mencakup pendekatan penyelesaian di luar pengadilan dan peran korban dalam proses hukum...	2025-10-03 00:00:00
3	Analisis Yuridis Sengketa Tanah Adat di Indonesia	Tulisan ini menjelaskan konsep tanah adat menurut UUPA dan peraturan turunannya, serta analisis terhadap beberapa kasus sengketa tanah adat yang diputus oleh Mahkamah Konstitusi...	2025-10-09 00:00:00
4	Penerapan Asas Legalitas dalam Hukum Pidana Indonesia	Artikel ini menguraikan makna asas legalitas menurut Pasal 1 KUHP dan prinsip nullum crimen sine lege. Ditekankan pentingnya asas ini untuk menjamin kepastian hukum dan mencegah penyalahgunaan kekuasaan oleh aparat penegak hukum...	2025-10-10 00:00:00

Gambar 7 Halaman Menu Lakukan Clustering

- Tampilan grafik yang terdapat pada menu lakukan clustering akan menampilkan diagram distribusi dokumen per cluster, gambar diagram seperti terlihat pada gambar 8.



Gambar 8 Halaman Grafik

- Klik menu cari dokumen yang terdapat pada menu utama admin kemudian sistem akan menampilkan hasil cluster dari setiap dokumen hukum, menu cari dokumen seperti terlihat pada gambar 9.

ID	Judul Dokumen	Cluster	File Dokumen	Nilai Similarity
3	Implementasi Restorative Justice dalam Penegakan Hukum Pidana di Indonesia	2	Tidak ada file	0.1369
10	Penerapan Asas Legalitas dalam Hukum Pidana Indonesia	2	Tidak ada file	0.1158
8	Tinjauan Hukum Pidana terhadap Kekerasan Berbasis Gender Online	1	Tidak ada file	0.0843
5	Analisis Hukum Perjanjian Kerja Berdasarkan Undang-Undang Cipta Kerja	1	Tidak ada file	0.0716
6	Perlindungan Hukum Terhadap Konsumen dalam Transaksi E-Commerce	1	Tidak ada file	0.0496
2	Kajian Hukum Perlindungan Data Pribadi Berdasarkan UU No. 27 Tahun 2022	0	Tidak ada file	0.0355

Gambar 9 Halaman Menu Cari Dokumen

SIMPULAN

Dari hasil pembahasan pada sebelumnya yang telah diuraikan, maka peneliti mengambil kesimpulan sebagai berikut:

- Untuk mengimplementasikan clustering dokumen hukum berdasarkan kesamaan isi teks. Proses ini diawali dengan tahap preprocessing dokumen untuk mengubah dokumen ke dalam bentuk vektor numerik. Hasil penerapan algoritma K-Means menunjukkan bahwa dokumen hukum dapat dikelompokkan ke dalam beberapa cluster yang merepresentasikan topik atau kategori hukum tertentu. Dengan demikian dokumen hukum yang memiliki kemiripan konteks dapat dikelompokkan secara otomatis tanpa proses manual sehingga mempermudah pengorganisasian arsip digital di kantor hukum.
- Hasil clustering dari algoritma K-Means terbukti dapat mengoptimalkan sistem pencarian dokumen hukum. Dengan adanya pengelompokan, sistem tidak lagi mencari dokumen di seluruh *database*, tetapi hanya di cluster yang relevan dengan kata kunci pencarian pengguna. Pendekatan ini mengurangi waktu pencarian, meningkatkan akurasi hasil pencarian, serta membuat proses penelusuran dokumen menjadi lebih terarah dan efisien. Sistem juga mampu menampilkan hasil pencarian dengan tingkat relevansi yang lebih tinggi karena proses pencarian dilakukan pada dokumen-dokumen dengan topik yang sama.
- Hasil pencarian yang lebih relevan dan cepat dengan pembagian cluster seperti cluster 0 dokumen Dokumen bertema hukum pidana, korupsi, dan keadilan restorative, cluster 1 Dokumen bertema perlindungan data pribadi, ITE, dan kekerasan berbasis gender online dan cluster 2 Dokumen bertema hukum perdata dan

administrasi, seperti perjanjian kerja, konsumen, dan tanah adat.

DAFTAR PUSTAKA

- Arif, A., Dwi Christyanti, R., & Kaltara, U. (2022). Clustering Calon Penerima Zakat Menggunakan Metode K-Means (Ratna Dwi Christyanti) (Studi Kasus di Provinsi Kalimantan Utara). *73 SMARTICS Journal*, 8(2), 73–79. <https://doi.org/10.21067/10.21067/smartics.v8i2.7531>
- Fahrul Fauji, L. F. (2025). *Penerapan Data Mining Menggunakan K-Means Clustering Dalam Mengelompokkan Tingkat Kesulitan Mata Pelajaran*. 6(3). <https://doi.org/10.47065/josh.v6i3.6959>
- Kusnadi, Y., & Putri, M. S. (2021). Clustering Menggunakan Metode K-Means Untuk Menentukan Prioritas Penerima Bantuan Bedah Rumah (Studi Kasus : Desa Ciomas Bogor). *Jurnal Teknologi Informatika Dan Komputer*, 7(1), 17–24. <https://doi.org/10.37012/jtik.v7i1.498>
- Maoulana, R., Irawan, B., & Bahtiar, A. (2024). Data Mining Dalam Konteks Transaksi Penjualan Hijab Dengan Menggunakan Algoritma Clustering K-Means. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 515–521. <https://doi.org/10.36040/jati.v8i1.850>
- Norshahlan, M., Jaya, H., & Kustini, R. (2023). Penerapan Metode Clustering Dengan Algoritma K-means Pada Pengelompokan Data Calon Siswa Baru. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, 2(6), 1042. <https://doi.org/10.53513/jursi.v2i6.9148>
- Nur Afidah, N. (2023). Penerapan Metode Clustering dengan Algoritma K-means untuk Pengelompokan Data Migrasi Penduduk Tiap Kecamatan di Kabupaten Rembang. *PRISMA, Prosiding Seminar Nasional Matematika*, 6, 729–738.
- Ramadana, F. D., Pratama, W. P., & Lingga, C. (2025). *Implementasi Algoritma K-Means Clustering terhadap Tingkat Kepuasan Peserta LKP Multi Talenta Komputer Purwokerto*. 3.