



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 4 No. 2 (2025) pp: 4227-4232

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Aplikasi Data Mining dalam Pengelompokan Informasi Multidisiplin

Fahri Patir Ramadhan¹, Muhammad Dzulfikar Apandi Putra², Muhammad Rafli Reinanda³, Naufal Akmal Maulana⁴, Zurnan Alfian⁵

¹Fakultas Ilmu Komputer, Prodi Teknik Informatika, Universitas Pamulang, Kota Tangerang Selatan, Indonesia.

Email: fahrifatirramadan099@gmail.com, mdzfy08@gmail.com, reinandarafli@gmail.com,
naufalakmalmaulana12@gmail.com

Abstrak

Tingginya volume literatur akademik dari berbagai disiplin ilmu menimbulkan tantangan dalam mengelola dan menganalisis informasi secara efektif. Penelitian ini mengangkat permasalahan kurangnya metode otomatisasi dalam pengelompokan konten multidisipliner. Sebagai solusinya, diterapkan algoritma *K-Means Clustering* untuk mengelompokkan data teks yang bersumber dari empat buku akademik lintas bidang: sosial-politik, keuangan pribadi, teknologi informasi, dan ekonomi makro. Metode yang digunakan mencakup ekstraksi manual teks penting, praproses data dengan *tokenisasi* dan *TF-IDF*, serta penerapan algoritma K-Means untuk membentuk kluster berdasarkan kesamaan tematik. Hasil penelitian menunjukkan bahwa metode ini berhasil mengelompokkan teks ke dalam kluster yang konsisten dengan bidang masing-masing, dengan nilai Silhouette Score sebesar 0,968 yang mengindikasikan kualitas pemisahan antar kelompok sangat baik. Temuan ini memperkuat efektivitas algoritma K-Means dalam analisis literatur multidisipliner dan menunjukkan potensinya sebagai alat bantu dalam eksplorasi konseptual dan klasifikasi konten akademik secara otomatis.

Kata kunci: Pengelompokan Data, K-Means, Literatur Akademik, Multidisipliner, Analisis Teks.

1. PENDAHULUAN

Seiring dengan pesatnya kemajuan teknologi informasi, ketersediaan data dari berbagai sumber—termasuk literatur ilmiah—mengalami pertumbuhan yang sangat signifikan. Peningkatan volume dan keragaman data ini menuntut adanya pendekatan analitis yang lebih efektif dan efisien untuk menggali informasi yang relevan dan bermakna. Dalam konteks ini, data mining menjadi salah satu metode yang banyak diminati, karena mampu mengidentifikasi pola tersembunyi dan relasi yang tidak tampak secara eksplisit dalam kumpulan data berskala besar dan kompleks.

Salah satu teknik dalam data mining yang banyak digunakan untuk tujuan eksploratif adalah *clustering*, yakni proses pengelompokan data berdasarkan kemiripan karakteristik tertentu tanpa memerlukan label atau kategori awal. Di antara berbagai algoritma clustering yang tersedia, *K-Means Clustering* menjadi salah satu metode yang paling populer karena kesederhanaan implementasinya serta kemampuannya dalam menangani data berskala besar dengan efisiensi tinggi. Meskipun teknik ini telah diterapkan luas dalam berbagai bidang seperti pemasaran, analisis pelanggan, dan segmentasi pasar, pemanfaatannya dalam domain literatur akademik, terutama yang melibatkan disiplin ilmu yang beragam, masih jarang dieksplorasi secara mendalam.

Penelitian ini berupaya menggali sejauh mana algoritma K-Means dapat dimanfaatkan untuk mengelompokkan konten dari buku-buku akademik yang berasal dari berbagai disiplin ilmu, termasuk namun tidak terbatas pada pembangunan sosial-politik, manajemen keuangan, teknologi informasi, dan ekonomi. Proses ini diawali dengan tahap ekstraksi manual konten penting dari masing-masing buku, kemudian dilanjutkan dengan praproses teks untuk menghasilkan representasi data yang dapat dianalisis secara komputasional. Dengan pendekatan ini, penelitian bertujuan tidak hanya untuk menguji efektivitas K-Means dalam mendeteksi kesamaan tematik di antara literatur multidisipliner, tetapi juga untuk mengeksplorasi kemungkinan terbentuknya kluster

yang merefleksikan keterkaitan konseptual lintas bidang keilmuan. Hasil yang diperoleh diharapkan dapat memberikan perspektif baru terhadap pemetaan pengetahuan dan mendorong pengembangan metode analisis teks yang lebih adaptif dalam konteks literatur akademik.

Pendekatan ini juga membuka peluang untuk mengidentifikasi kecenderungan topik yang saling bersinggungan antar disiplin, sehingga dapat memperkuat pemahaman terhadap dinamika interdisipliner dalam dunia akademik. Dengan mengelompokkan konten berdasarkan kemiripan tematik menggunakan K-Means, penelitian ini tidak hanya berkontribusi pada pengembangan metode analisis teks dalam ranah ilmu perpustakaan dan informasi, tetapi juga berpotensi menjadi alat bantu yang efektif dalam proses kurasi, rekomendasi, atau klasifikasi literatur akademik secara otomatis. Pada akhirnya, integrasi antara teknik data mining dan studi literatur diharapkan mampu memperluas cakrawala kajian ilmiah melalui pendekatan yang lebih kuantitatif, sistematis, dan berbasis data.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan sifat eksploratif, dengan tujuan menguji kemampuan algoritma *K-Means Clustering* dalam mengelompokkan dokumen teks dari berbagai bidang keilmuan. Fokus utama berada pada pengorganisasian literatur akademik secara otomatis ke dalam kelompok yang memiliki kedekatan tematik berdasarkan isi konten. Pendekatan ini melibatkan tahapan yang sistematis, mulai dari pengumpulan sumber data, transformasi teks menjadi bentuk numerik, penerapan algoritma klusterisasi, hingga evaluasi hasil kelompok yang terbentuk.

2. 1. Pengumpulan Data

Sumber data diperoleh dari empat buku akademik yang masing-masing mewakili disiplin ilmu yang berbeda. Pemilihan literatur dilakukan secara purposif berdasarkan keberagaman bidang, sehingga mampu mencerminkan variasi tematik. Keempat buku tersebut meliputi:

- *Pembangunan, Kebebasan, dan “Mukjizat” Orde Baru*
- *Cash Flow Management untuk Orang Awam & Pemula*
- *Data Mining*
- *Pengantar Ekonomi Makro Edisi ke-7*

2. 2. Penerapan Algoritma K-Means

Setelah data diproses, algoritma K-Means diterapkan dengan menentukan jumlah kluster (k) sebanyak 4, sesuai jumlah Langkah berikutnya adalah transformasi data teks menjadi bentuk numerik menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF)*, yang bertujuan untuk menangkap pentingnya kata dalam konteks dokumen dan keseluruhan korpus. Setelah transformasi dilakukan, diterapkan algoritma K-Means dengan jumlah kluster (k) ditentukan sebanyak empat, disesuaikan dengan jumlah domain buku yang dianalisis.

Model K-Means bekerja dengan memetakan dokumen ke dalam ruang vektor multidimensi, dan membagi data berdasarkan jarak terdekat dari titik pusat (centroid) masing-masing kluster. Proses pelatihan ini bersifat *unsupervised*, tanpa adanya label kelas sebelumnya.

2. 3. Analisis Hasil

Evaluasi dilakukan melalui dua pendekatan: analisis kualitatif dan kuantitatif. Dari sisi kualitatif, setiap kluster yang terbentuk dianalisis isinya untuk melihat apakah dokumen-dokumen di dalamnya menunjukkan keterkaitan topik yang konsisten dan logis. Sementara itu, dari aspek kuantitatif digunakan metrik *Silhouette Score*—sebuah indikator statistik yang mengukur seberapa baik objek cocok dengan kluster-nya dibandingkan dengan kluster lain. Nilai Silhouette Score mendekati 1 menunjukkan bahwa data memiliki pemisahan yang jelas dan pengelompokan yang optimal. Dalam penelitian ini, skor yang diperoleh sangat tinggi, mencerminkan efektivitas algoritma dalam membedakan konten berdasarkan bidang keilmuan.

3. PEMBAHASAN

a. Pembangunan, Kebebasan, dan “Mukjizat” Orde Baru

Pengumpulan data dilakukan dengan menyeleksi kutipan penting dari isi buku yang membahas dinamika politik dan sosial selama masa pemerintahan Orde Baru. Potongan teks yang dipilih kemudian diolah melalui tahapan praproses yang mencakup pemisahan kata (tokenisasi), penghapusan kata-kata umum (stopwords), dan transformasi ke dalam bentuk vektor menggunakan teknik *Term Frequency–Inverse Document Frequency (TF-IDF)*. Data vektor ini selanjutnya dianalisis menggunakan algoritma K-Means untuk menemukan pola kemiripan tematik di antara bagian-bagian teks yang dianalisis.

Kumpulan data yang digunakan terdiri atas teks naratif yang memuat tema seputar pembangunan nasional, kontrol politik, kebebasan sipil, serta kebijakan negara pada era Orde Baru. Istilah-istilah seperti “kestabilan nasional,” “represi politik,” dan “orde pembangunan” menjadi indikator dominan dalam representasi tematik dataset. Teks-teks tersebut secara eksplisit mencerminkan wacana politik dan sosiologis yang relevan dengan konteks sejarah Indonesia.

Proses klusterisasi berhasil memisahkan konten ini ke dalam satu kelompok yang secara tematik konsisten dengan isu sosial-politik. Hasil ini menunjukkan bahwa algoritma K-Means mampu menangkap karakteristik semantik yang khas dalam literatur politik dan membedakannya secara jelas dari dokumen yang berasal dari disiplin lain. Keberhasilan ini memperkuat potensi algoritma ini dalam mengidentifikasi dan mengelompokkan dokumen berdasarkan kedekatan tema, bahkan dalam konteks teks yang bersifat naratif dan kualitatif seperti studi sejarah politik.

b. Cash Flow Management untuk Orang Awam & Pemula

Tahun/ Bulan	Bahan Makana n	Makanan jadi, Minuman , Rokok, Dan Tembaka u	Perumaha n air, Listrik, Gas, dan Bahan Bakar	Sandan g	Kesehata n	Pendidikan , Rekreasi dan Olahraga	Transpor, Komunikas i dan Jasa Keuangan	Indek s umum
2006	12.94	6.36	4.83	6.84	5.87	8.13	1.02	6.60
Desember	3.12	1.11	0.74	0.13	1.05	0.07	0.10	1.21
Novembe r	0.64	0.47	0.20	0.70	0.42	0.03	-0.21	0.34
Oktober	2.17	0.64	0.26	1.00	0.29	0.10	0.46	0.86
Septembe r	0.62	0.13	0.28	-0.13	0.31	1.84	-0.01	0.38
Agustus	-0.34	0.35	0.30	0.35	0.33	4.77	0.01	0.33
Juli	0.99	0.31	0.21	0.36	0.06	0.69	0.08	0.45
Juni	1.12	0.26	0.32	-0.08	0.27	0.25	0.10	0.45
Mei	0.28	0.30	0.30	2.03	0.57	0.07	0.17	0.37
April	-0.85	0.43	0.42	0.70	0.58	0.09	0.07	0.05
Maret	-0.88	0.58	0.36	0.15	0.39	0.12	0.13	0.03
Febuari	1.18	0.65	0.55	0.72	0.40	-0.28	0.16	0.58
Januari	4.29	0.94	0.70	0.73	1.06	0.20	-0.05	1.36

Analisis dimulai dengan mengambil bagian-bagian utama dari buku yang membahas konsep dasar pengelolaan keuangan pribadi, khususnya bagi individu non-akuntan. Teks-teks penting yang menjelaskan alur keuangan rumah tangga, seperti perencanaan anggaran, pengaturan pengeluaran, dan alokasi tabungan, dipilih secara manual. Proses praproses meliputi penghapusan tanda baca, tokenisasi kata, normalisasi huruf, serta penghapusan

stopwords. Representasi teks kemudian diubah menjadi bentuk numerik menggunakan pendekatan TF-IDF, dan diterapkan algoritma K-Means untuk menemukan struktur kluster berdasarkan kesamaan istilah.

Analisis dimulai dengan mengambil bagian-bagian utama dari buku yang membahas konsep dasar pengelolaan keuangan pribadi, khususnya bagi individu non-akuntan. Teks-teks penting yang menjelaskan alur keuangan rumah tangga, seperti perencanaan anggaran, pengaturan pengeluaran, dan alokasi tabungan, dipilih secara manual. Proses praproses meliputi penghapusan tanda baca, tokenisasi kata, normalisasi huruf, serta penghapusan stopwords. Representasi teks kemudian diubah menjadi bentuk numerik menggunakan pendekatan TF-IDF, dan diterapkan algoritma K-Means untuk menemukan struktur kluster berdasarkan kesamaan istilah.

Hasil pengelompokan memperlihatkan bahwa teks dari buku ini membentuk kluster tersendiri yang tidak tumpang tindih dengan kluster dari literatur akademik lain, seperti teknologi atau ekonomi makro. Algoritma K-Means berhasil mengenali pola terminologi yang lebih praktis dan berorientasi pada keseharian finansial. Temuan ini menunjukkan bahwa pendekatan klusterisasi mampu mengisolasi konten berdasarkan konteks penggunaan dan tujuan komunikasi—dalam hal ini, edukasi finansial non-formal. Hasil klusterisasi juga memberikan peluang untuk otomatisasi klasifikasi dokumen edukasi berbasis praktik keuangan dalam aplikasi sistem rekomendasi literatur atau e-learning.

c. Data Mining

Kerawat Desa	Luas (Hektar)
Kepala Desa	6,80
Sekretaris Desa	3,50
Kepala Urusan Kersa I	1,30
Kepala Urusan Kersa II	0,70
Kepala Urusan Keuangan	1,40
Kepala Urusan Pembangunan	0,85
Kepala Urusan Pemerintahan	1,50
Kepala Urusan Umum	1,45
Kepala Dusun (5) : 5 x 1,40	7,00
Jumlah	24,50

Analisis dilakukan dengan menyeleksi bagian-bagian penting dari buku yang membahas teori dasar dan praktik implementasi data mining, terutama algoritma *clustering* dan teknik pra-analisis data. Kutipan teoritis yang menjelaskan tahapan preprocessing, metode analitik, serta penerapan algoritma dipilih secara manual. Selanjutnya, dilakukan pembersihan teks dan konversi ke dalam format numerik menggunakan pendekatan *TF-IDF*. Representasi vektor dari teks kemudian menjadi input untuk algoritma K-Means guna menemukan kesamaan tematik dalam konten..

Korpus data mencakup istilah teknis dan metodologis seperti “clustering,” “preprocessing,” “atribut numerik,” “algoritma,” serta “transformasi data.” Sumber teks juga mencakup ilustrasi penggunaan data dalam konteks administratif, seperti tabel distribusi tugas perangkat desa, yang memberikan kombinasi antara teori dan data berbasis aplikasi. Konten ini secara dominan mencerminkan literatur dari bidang teknik informatika dan sistem informasi.

Hasil dari klusterisasi menunjukkan bahwa dokumen ini tergolong dalam kluster yang sangat spesifik terhadap tema teknologi dan komputasi. Teks dari buku ini dapat dibedakan secara jelas dari literatur lain yang lebih bersifat ekonomi atau sosial karena karakteristik terminologinya yang khas. Keberhasilan algoritma K-Means dalam mengelompokkan dokumen ini ke dalam kluster teknis menegaskan efektivitas metode ini dalam memetakan

konten berbasis konteks disiplin ilmu, serta membuka peluang untuk mengotomatisasi klasifikasi teks berbasis teknologi di berbagai aplikasi sistem analisis data atau perpustakaan digital.

d. Buku Pengantar Ekonomi Makro

Negara	Periode	PDB Riil Tiap Penduduk Pada Awal Periode ^a	PDB Riil Tiap Penduduk Pada Akhir Periode ^a	Angka Pertumbuhan (Pertahun)
Jepang	1890-2010	\$1.517	\$34.810	2,65%
Brasil	1900-2010	785	10.980	2,43
Meksiko	1900-2010	1.169	14.350	2,31
Tiongkok	1900-2010	723	7.520	2,15
Jerman	1870-2010	2.204	38.410	2,06
Kanada	1870-2010	2.397	38.370	2,00
Amerika Serikat	1870-2010	4.044	47.210	1,77
Argentina	1900-2010	2.314	15.470	1,74
India	1900-2010	681	3.330	1,45
Inggris	1870-2010	4.853	35.620	1,43
Indonesia	1900-2010	899	4,180	1,41
Pakistan	1900-2010	744	2.760	1,20
Bangladesh	1900-2010	629	1.800	0,096

Proses analisis dimulai dengan memilih bagian-bagian utama dari buku yang membahas topik ekonomi makro, khususnya materi mengenai pertumbuhan ekonomi, produk domestik bruto (PDB), kebijakan moneter, inflasi, serta indikator ekonomi lainnya. Kutipan-kutipan ini kemudian diproses menggunakan teknik pra-proses teks, yang mencakup pemisahan kata (tokenisasi), pembersihan kata tidak bermakna (stopword removal), serta perubahan format teks menjadi vektor numerik melalui metode *Term Frequency–Inverse Document Frequency (TF-IDF)*. Hasil representasi numerik tersebut kemudian dianalisis dengan algoritma K-Means untuk membentuk klaster berdasarkan pola istilah yang serupa.

Data terdiri dari teks yang mengandung konsep-konsep makroekonomi, seperti “pertumbuhan jangka panjang,” “pendapatan per kapita,” “kebijakan fiskal,” dan “permintaan agregat.” Selain itu, tabel statistik yang menjelaskan perbandingan pertumbuhan ekonomi antar negara dan periode juga diolah menjadi bagian dari representasi numerik. Kumpulan data ini sangat representatif dalam menggambarkan struktur dan terminologi yang umum ditemukan dalam literatur ekonomi akademik tingkat dasar hingga menengah..

Hasil dari proses klasterisasi menunjukkan bahwa data dari buku ini tergabung dalam satu kelompok yang secara konsisten merepresentasikan wacana ekonomi makro. K-Means mampu memisahkan konten ini dari teks lain, seperti yang membahas keuangan praktis atau ilmu komputer, karena adanya perbedaan dalam struktur istilah dan cakupan tematik. Hal ini membuktikan bahwa algoritma mampu mengidentifikasi kerangka semantik yang kompleks dalam konteks ekonomi dan mengelompokkannya secara efisien. Lebih jauh, pendekatan ini berpotensi diterapkan dalam sistem klasifikasi otomatis untuk buku ajar ekonomi, baik dalam skala perpustakaan digital maupun platform e-learning.

KESIMPULAN

Penelitian ini membuktikan bahwa pendekatan *unsupervised learning* melalui algoritma K-Means dapat digunakan secara efektif untuk mengelompokkan teks dari berbagai bidang akademik ke dalam klaster yang bermakna secara tematik. Dengan melalui serangkaian tahapan seperti ekstraksi teks, praproses menggunakan metode TF-IDF, serta normalisasi numerik, sistem berhasil membentuk pengelompokan yang mencerminkan karakteristik masing-masing disiplin ilmu secara jelas. Evaluasi kuantitatif menggunakan *Silhouette Score* menghasilkan nilai sebesar 0,968, yang mengindikasikan bahwa klaster-klaster yang terbentuk memiliki pemisahan yang sangat baik dan tingkat keseragaman internal yang tinggi. Selain itu, hasil visualisasi juga menunjukkan pola pengelompokan yang dapat ditafsirkan dengan mudah, menambah dimensi interpretatif dalam analisis data teks. Secara keseluruhan, studi ini menunjukkan bahwa integrasi teknik *clustering* berbasis data mining dapat dimanfaatkan sebagai metode analitik yang efektif untuk menyusun, mengeksplorasi, dan memahami struktur konseptual dalam literatur akademik yang bersifat multidisipliner. Temuan ini membuka potensi pemanfaatan lebih luas dalam sistem klasifikasi otomatis, pemetaan pengetahuan, hingga pengembangan alat bantu bagi proses pembelajaran dan riset di era digital.

REFERENCES

- Thee, K. W. (2004). *Pembangunan, kebebasan dan "mukjizat" Orde Baru*. Jakarta: Buku Kompas.
- Mankiw, N. G. (2018). *Pengantar ekonomi makro*. Jakarta: Salemba Empat.
- Hermawati, F. A. (2009). *Data mining*. Yogyakarta: Penerbit Andi.
- Ariefiansyah, M. (2012). *Cash flow management untuk orang awam & pemula*. Bekasi: Laskar Aksara.