

SKRIPSI

**IMPLEMENTASI K-MEANS CLUSTERING DALAM MENGELOMPOKKAN
MINAT BACA MAHASISWA DIPERPUSTAKAAN
(STUDI KASUS : UNIVERSITAS PERADABAN)**



Diajukan Untuk Memenuhi Salah Satu Syarat Guna
Memperoleh Gelar Sarjana Komputer

Oleh:
Fitri Yanah
42420014

**JURUSAN INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS PERADABAN
2025**

SKRIPSI

**IMPLEMENTASI K-MEANS CLUSTERING DALAM MENGELOMPOKKAN
MINAT BACA MAHASISWA DIPERPUSTAKAAN
(STUDI KASUS : UNIVERSITAS PERADABAN)**



Diajukan Untuk Memenuhi Salah Satu Syarat Guna
Memperoleh Gelar Sarjana Komputer

Oleh:
Fitri Yanah
42420014

**JURUSAN INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS PERADABAN
2025**

PERSETUJUAN SKRIPSI

JUDUL : IMPLEMENTASI K-MEANS CLUSTERING DALAM
MENGELOMPOKKAN MINAT BACA MAHASISWA DI
PERPUSTAKAAN
(STUDI KASUS : UNIVERSITAS PERADABAN)
NAMA : FITRI YANAH
NIM : 42420014

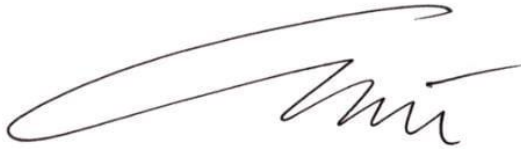
Skripsi ini telah diperiksa dan disetujui oleh Dosen pembimbing tugas akhir guna
mencapai Gelar Sarjana Komputer Strata satu pada Fakultas Sains dan
Teknologi, Universitas Peradaban

Bumiayu, 21 Juni 2025

Mengetahui,

Pembimbing I,

Pembimbing II,



Khurotul Aeni, M.Kom

NIDN.0618098802



Fathulloh, S.T., M.Kom

NIDN.0623048102

Ketua Jurusan,



Khurotul Aeni, M.Kom

NIDN.0618098802

PENGESAHAN SKRIPSI

JUDUL : IMPLEMENTASI K-MEANS CLUSTERING DALAM
MENGELOMPOKKAN MINAT BACA MAHASISWA DI
PERPUSTAKAAN
(STUDI KASUS : UNIVERSITAS PERADABAN)

NAMA : FITRI YANAH

NIM : 42420014

Skripsi ini telah diujikan dan dipertahankan dihadapan Dewan Penguji Skripsi.
Pada Tanggal 26 Juli 2025. Menurut pandangan kami. Skripsi ini memadai dari
segi kualitas untuk memperoleh gelar Sarjana Komputer (S.Kom).

Bumiayu, 06 September 2025

Nama Penguji

1. Nurul Mega Saraswati, M.Kom
2. Sorikhi, M.Kom
3. Khurotul Aeni, M.Kom
4. Fathulloh, S.T., M.Kom

Tanda Tangan

Dekan Fakultas Sains dan Teknologi
Universitas Peradaban

Rizki Ngor. Prasetyono, M.Pd.
NIDN. 0611099101

Ketua Jurusan
Informatika

Nurul Mega Saraswati, M.Kom
NIDN. 0606069102

PERNYATAN ORISINILITAS

JUDUL IMPLEMENTASI K-MEANS CLUSTERING DALAM
 MENGELOMPOKKAN MINAT BACA MAHASISWA DI
 PERPUSTAKAAN
 (STUDI KASUS : UNIVERSITAS PERADABAN)

NAMA FITRI YANAH

NIM 42420014

“Saya menyatakan dan bertanggung jawab dengan sebenar-benarnya bahwa skripsi ini adalah hasil karya saya sendiri dan bukan merupakan pengambilan tulisan atau pikiran orang lain yang saya akui sebagai tulisan dan pikiran saya, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka. Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa skripsi ini hasil plagiat maka saya bersedia menerima sanksi berupa pencabutan gelar sarjana computer beserta segala hak dan kewajiban yang melekat pada gelar tersebut”.

Bumiayu, 26 Juli 2025

Penulis

Fitri Yanah

NIM. 42420014.

ABSTRACT

University libraries face challenges in acquiring relevant book collections due to difficulties in efficiently identifying students' reading interests. This study aims to implement text mining techniques using the K-Means clustering algorithm to group students' reading interests based on library book borrowing data. A quantitative approach was used in this research, utilizing 1,524 secondary data records of book borrowing history. Book titles were processed through several preprocessing stages (case folding, tokenization, stopword removal, and stemming) and word weighting using TF-IDF, then grouped using the K-Means algorithm. The clustering results were evaluated using the Davies-Bouldin Index (DBI), which indicated that the optimal number of clusters was two, with the lowest DBI score of 2.5853. The clustering results successfully identified two categories of reading interests: research methodology and financial management, and introduction to social sciences and basic management. It is concluded that the K-Means algorithm is effective for mapping students' reading interests, thus providing data-driven recommendations for libraries in developing book collections strategically.

Keywords: Reading interest, Text Mining, K-Means Clustering, Library, Davies-Bouldin Index.

ABSTRAK

Perpustakaan perguruan tinggi menghadapi tantangan dalam pengadaan koleksi buku yang relevan karena kesulitan mengidentifikasi minat baca mahasiswa secara efisien. Penelitian ini bertujuan untuk mengimplementasikan teknik *text mining* menggunakan *algortima k-means clustering* untuk mengelompokkan minat baca mahasiswa berdasarkan data peminjaman buku di perpustakaan. Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan 1524 data sekunder riwayat peminjaman buku. Data judul buku diolah melalui tahapan *preprocessing* (*case folding, tokenize, stopword removal, stemming*) dan pembobotan kata menggunakan *TF-IDF*, kemudian dikelompokkan dengan *algortima k-means*. Evaluasi hasil klasterisasi menggunakan *Davies-Bouldin Index* (DBI) menunjukkan bahwa jumlah klaster terbaik adalah 2 klaster dengan nilai DBI terendah sebesar 2.5853. Hasil pengelompokkan berhasil mengidentifikasi dua kategori minat baca, yaitu metodologi penelitian dan manajemen keuangan, pengantar ilmu sosial dan manajemen dasar. Disimpulkan bahwa algoritma *K-Means* efektif untuk memetakan minat baca mahasiswa, sehingga dapat memberikan rekomendasi berbasis data bagi perpustakaan dalam strategis pengembangan koleksi buku.

Kata Kunci: Minat baca, *Text Mining*, *K-Means Clustering*, Perpustakaan, *Davies-Bouldin Index*.

KATA PENGANTAR

Bismillahirrahmaanirrahiim

Alhamdulillahirobbil'aalamin, Segala puji syukur bagi Allah ‘azza wa jalla dengan segala rahmat, Nikmat, Hidayah dan Inaya-Nya, Sehingga penulis dapat menyelesaikan skripsi ini. Shalawat dan salam semoga tetap tercurahkan dan terlimpahkan kepada Sang Baginda Rasul *Muhammad SAW*, berserta kepada keluarga, para sahanat dan penerus risalahnya, karena atas segala perjuangan beliau selama hidup telah mewariskan ilmu berserta penuntun hidup yang mencerahkan umat manusia, semoga kita sebagai penerus risalah beliau, selalu mendapatkan syafa'atnya Amiin.

Alhamdulillah, dengan segala ikhtiar dan doa, penulis dapat menyelesaikan penyusunan skripsi dengan judul: *Implementasi K-means Clustering Dalam Mengelompokkan Minat Baca Mahasiswa Di Perpustakaan (Studi Kasus: Universitas Peradaban)* untuk dijadikan syarat memperoleh gelar Sarjana Komputer di skripsi ini tentu tidak akan selesai tanpa adanya bantuan dan kerjasama dari berbagai pihak. Oleh karena itu, melalui kesempatan ini penulis selayaknya menyampaikan terima kasih sebesar-besarnya kepada:

1. Bapak Dr. Muh. Kadarisma, S.H., M.Si. Selaku Rektor Universitas Peradaban Bumiayu
2. Bapak Rizki Noor Prasetyono, M.Pd. Selaku Dekan Fakultas Sains dan Teknologi Universitas Peradaban Bumiayu
3. Ibu Khurotul Aeni, M.Kom. Selaku Kepala Program Studi Informatika
4. Ibu Khurotul Aeni, M.Kom. Selaku Dosen Pembimbing Skripsi yang selalu sabar dalam memberikan arahan supaya hasil dari karya ilmiah ini maksimal serta selalu mengingatkan tentang penyusunan skripsi.
5. Bapak Fathulloh, S.T., M.Kom. Selaku Dosen Pembimbing Skripsi yang sabar dan memberikan arahan supaya hasil dari karya ilmiah ini maksimal.
6. Seluruh Dosen Informatika Universitas Peradaban Bumiayu.
7. Kedua orang tua saya, yang selalu memberikan doa dan dukungan penuh, mereka adalah alasan saya untuk tetap bertahan sejauh ini.

8. Adik tercinta saya, yang selalu memberikan doa dan semangat kepada saya.
9. Kaka saya tercinta, yang selalu mendoakan serta memberi semangat kepada saya agar selalu kuat.
10. Trian Efendi, selaku orang yang selalu menemani dan memberikan semangat dan mendengarkan keluh kesah penulis.
11. Teman-teman kelas, terimakasih atas dukungan dan kebersamaan selama penulis menempuh studi di program studi S-1 informatika, Fakultas Sains dan Teknologi, Universitas Peradaban Bumiayu.
12. Tidak lupa untuk semua pihak yang memberikan penulis dukungan yang tidak dapat disebutkan satu persatu. Terimakasih banyak.

Penulis menyadari bahwa masih banyak kekurangan dari skripsi ini. Oleh karena itu, penulis memohon maaf atas segala kekurangan yang terjadi selama proses penyusunan skripsi ini.

Kepada semua yang telah memberikan ilmu dan pengalaman kepada penulis, semoga ilmu yang diberikan dapat bermanfaat bagi penulis sampai pada masa yang akan datang, semoga Allah SWT membalas kebaikan semuanya. Dengan selesainya skripsi ini penulis berharap semoga dapat bermanfaat bagi penulis dan pembaca khususnya. Amiin.

Bumiayu, 26 Juli 2025

Penulis



Fitri Yanah

NIM. 42420014

DAFTAR ISI

HALAMAN JUDUL.....	i
PERSETUJUAN SKRIPSI	ii
PENGESAHAN SKRIPSI	iii
PERNYATAN ORISINILITAS.....	iv
ABSTRACT.....	v
ABSTRAK	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Penelitian	4
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Peneliti Terkait	5
2.2 Landasan Teori	10
2.2.1 <i>Data Mining</i>	10
2.2.2 <i>Text mining</i>	12
2.2.3 <i>Text Preprocessing</i>	13
2.2.4 <i>TF-IDF</i>	13
2.2.5 Minat Baca	14
2.2.6 Perpustakaan	14
2.2.7 <i>K-Means Clustering</i>	14
2.2.8 <i>Davies Bouldin Index (DBI)</i>	16
2.2.9 <i>Python</i>	17

2.2.10	<i>Library</i>	17
2.2.11	<i>Pandas</i>	18
2.2.12	<i>Numpy</i>	18
2.2.13	<i>Flask</i>	18
2.2.14	<i>Matplotlib</i>	18
2.2.15	<i>Anaconda Navigator</i>	18
2.2.16	<i>Visual studio code</i>	19
2.3	Kerangka Pemikiran	19
BAB III METODE PENELITIAN.....		21
3.1	Tahapan Penelitian	21
3.2	Pengolahan Data.....	22
3.2.1	<i>Preprocessing Data</i>	24
3.2.2	<i>TF-IDF</i>	25
3.2.3	Metode Yang Diusulkan	25
3.2.4	Evaluasi	26
3.3	Perangkat Yang Digunakan.....	26
3.3.1	Perangkat keras (<i>Hardware</i>)	26
3.3.2	Perangkat lunak (<i>Software</i>)	26
BAB IV HASIL DAN PEMBAHASAN		27
4.1	Hasil Pengumpulan Data	27
4.2	Pengolahan Data.....	28
4.2.1	<i>Preprocessing Data</i>	28
4.2.2	<i>TF-IDF</i>	32
4.2.3	Implementasi <i>K-Means Clustering</i>	34
4.2.4	Evaluasi	39
4.3	Implementasi Program	40
4.3.1	Tampilan <i>GUI</i>	40
4.3.2	Tampilan Hasil <i>Preprocessing Data</i>	41
4.3.3	Tampilan Hasil <i>Clustering</i>	43
4.3.4	Hasil <i>WordCloud</i>	45
4.3.5	Tampilan Hasil Evaluasi Model.....	46
BAB V KESIMPULAN		48

5.1	Kesimpulan.....	48
5.2	Saran.....	48
DAFTAR PUSTAKA		49
LAMPIRAN.....		51

DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait	7
Tabel 2. 2 Penelitian Terkait Lanjutan I.....	8
Tabel 2. 3 Penelitian Terkait Lanjutan II	9
Tabel 4. 1 Peminjaman Buku.	27
Tabel 4. 2 Judul Buku.	28
Tabel 4. 3 <i>Case Folding</i>	29
Tabel 4. 4 <i>Tokenizing</i>	30
Tabel 4. 5 <i>Stopword Removal</i>	30
Tabel 4. 6 <i>Stemming</i>	31
Tabel 4. 7 Hasil <i>Preprocessing</i>	32
Tabel 4. 8 Contoh Judul buku.	33
Tabel 4. 9 <i>TF-IDF</i>	33
Tabel 4. 10 Contoh Dokumen Judul Buku.....	34
Tabel 4. 11 <i>Matriks Term</i> Dokumen.	35
Tabel 4. 12 Hasil Pengelompokan iterasi pertama.....	38
Tabel 4. 13 Hasil Nilai <i>DBI</i>	39
Tabel 4. 14 Penjelasan hasil nilai <i>DBI</i>	39

DAFTAR GAMBAR

Gambar 2. 1 Kerangka Pemikiran	19
Gambar 3. 1 Tahapan Penelitian	21
Gambar 3. 2 Pengolahan Data.....	23
Gambar 3. 3 <i>Preprocessing</i> Data	24
Gambar 3. 4 Tahapan <i>TF-IDF</i>	25
Gambar 3. 5 Tahapan <i>Algoritma K-means Clustering</i>	25
Gambar 4. 1 Tampilan Upload Dataset.....	41
Gambar 4. 2 Tampilan Hasil <i>Casefolding</i>	41
Gambar 4. 3 Tampilan Hasil <i>Tokenizing</i>	42
Gambar 4. 4 Tampilan Hasil <i>Stopword Removal</i>	42
Gambar 4. 5 Tampilan Hasil <i>Stemming</i>	43
Gambar 4. 6 Tampilan Hasil <i>TF-IDF</i>	43
Gambar 4. 7 Masukan Jumlah <i>Cluster</i>	44
Gambar 4. 8 Tampilan Hasil Model.....	44
Gambar 4. 9 Lanjutan Tampilan Hasil Model	44
Gambar 4. 10 Hasil <i>Wordcloud Cluster 0</i>	45
Gambar 4. 11 Hasil <i>Wordcloud Cluster 1</i>	46
Gambar 4. 13 Tampilan Hasil Evaluasi Model.....	46

DAFTAR LAMPIRAN

Lampiran 1 Observasi Perpustakaan.....	51
Lampiran 2 Surat Penelitian Perpustakaan.	52
Lampiran 3 Dokumentasi Wawancara.	53
Lampiran 4 Turnitin	54