



**KLASTERING BUKU PADA PERPUSTAKAAN MENGGUNAKAN
K-MEANS**

**TUGAS AKHIR
SKRIPSI**

**BENNY DANIEL BAHARI
NIM. 2010511083**

**PROGRAM STUDI S1 INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA
2026**

KATA PENGANTAR

Puji dan Syukur dipanjatkan kehadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan Tugas Akhir ini dengan baik. Tugas Akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom) pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.

Penyelesaian Tugas Akhir ini tidak terlepas dari kontribusi berbagai pihak. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang tulus kepada:

1. Tuhan Yang Maha Esa atas rahmat dan karunia-Nya yang tidak terhingga.
2. Kedua orang tua yang selalu memberikan dukungan agar penulis semangat menyelesaikan tugas akhir perkuliahan.
3. Bapak Prof, Dr. Ir. Supriyanto, selaku Dekan Fakultas Ilmu Komputer UPN Veteran Jakarta
4. Bapak Dr. Ridwan Raafi'udin, M.Kom, selaku Ketua Program Studi S1 Informatika.
5. Ibu Iin Ernawati, S.Kom., M.Si, selaku Dosen Pembimbing 1.
6. Ibu Kharisma Wiati Gusti, M.T, selaku Dosen Pembimbing 2.
7. Teman-teman baik di lingkungan Fakultas Ilmu Komputer UPN Veteran Jakarta maupun di luar lingkungan.
8. Seluruh pihak yang tidak dapat disebutkan satu – persatu.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang bersifat membangun dari berbagai pihak.

Jakarta, 16 April 2026



Benny Daniel Bahari

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri dan semua sumber yang dikutip maupun yang dirujuk telah dinyatakan dengan benar:

Nama : Benny Daniel Bahari

NIM : 2010511083

Program Studi : S1 Informatika

Bilamana pada kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia untuk bertanggung jawab dengan ketentuan yang berlaku.

Jakarta, 27 Juli 2026



Benny Daniel Bahari

**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMIS**

Sebagai civitas akademik Universitas Pembangunan Nasional “Veteran” Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Benny Daniel Bahari
NIM : 2010511083
Fakultas : Fakultas Ilmu Komputer
Program Studi : S1 Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non eksklusif (*Non - exclusive Royalty Free Right*) atas skripsi saya yang berjudul :

Klasifikasi Buku Pada Perpustakaan Menggunakan K-Means

Dengan Hak Bebas Royalti ini, Universitas Pembangunan Nasional “Veteran” Jakarta berhak menyimpan, mengalih media atau memformatkan, mengelola dalam bentuk pangkalan data (basis data), merawat dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 27 Juli 2026
Yang menyatakan,



Benny Daniel Bahari

LEMBAR PENGESAHAN

Judul : Klasifikasi Buku Pada Perpustakaan Menggunakan K-Means
Nama : Benny Daniel Bahari
NIM : 2010511083
Program Studi : S1 Informatika

Disetujui oleh :

Penguji 1:

Dr. Indra Permana Solihin, S.Kom, M.Kom



Penguji 2:

Nindy Irzavika, S.SI., M.T



Pembimbing 1:

Iin Ernawati, S.Kom., M.Si



Pembimbing 2:

Kharisma Wiati Gusti, M.T.

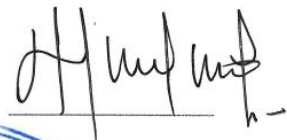


Diketahui oleh:

Koordinator Program Studi:

Dr. Ridwan Raafi'udin, M.Kom

NIP. 198805022021211001



Dekan Fakultas Ilmu Komputer

Prof. Dr. Ir. Supriyanto, S.T., M.Sc., IPM

NIP. 197605082003121002




Tanggal Ujian Tugas Akhir:

25 Mei 2026

ABSTRAK

Pengelolaan koleksi buku pada perpustakaan modern memerlukan sistem yang efisien untuk memudahkan pencarian dan pengelompokan buku. Perpustakaan Universitas Pembangunan Nasional Veteran Jakarta menghadapi kendala dalam mengorganisir ribuan judul buku dengan berbagai kategori. Penelitian ini bertujuan untuk melakukan klasterisasi judul buku menggunakan algoritma K-Means serta menerapkan teknik text mining. Data yang digunakan berupa 5715 data peminjaman buku. Proses penelitian mencakup tahap pra-pemrosesan seperti *case folding*, *tokenization*, normalisasi, *stopword removal* dan *stemming*. Kemudian, penelitian dilanjutkan dengan ekstraksi fitur menggunakan TF-IDF untuk mengubah data teks menjadi bentuk numerik. Proses klasterisasi dilakukan dengan K-Means dan dievaluasi menggunakan Elbow Method serta Davies-Bouldin Index (DBI). Hasil penelitian menunjukkan bahwa jumlah *cluster* optimal adalah $k = 5$ dengan nilai DBI sebesar 2.103. Metode K-Means mampu mengelompokkan judul buku berdasarkan kemiripan topik secara efektif. Hasil ini diharapkan dapat membantu pengelolaan koleksi perpustakaan serta memudahkan pengguna dalam menemukan buku yang relevan.

Kata kunci: *text mining, clustering, K-Means, TF-IDF, Davies-Bouldin Index*

ABSTRACT

Management of book collections in modern libraries requires an efficient system to facilitate searching and grouping processes. Library of Universitas Pembangunan Nasional Veteran Jakarta faces challenges in organizing thousands of book titles across various categories. This study aims to cluster book titles using the K-Means algorithm and apply text mining techniques. The data used consists of 5,715 book borrowing records. The research process includes preprocessing stages such as case folding, tokenization, normalization, stopword removal, and stemming, followed by feature extraction using TF-IDF to convert text data into numerical form. Clustering process is carried out using K-Means and evaluated using the Elbow Method and Davies-Bouldin Index (DBI). Results show that the optimal number of clusters is $k = 5$ with a DBI value of 2.103. K-Means method is able to effectively group book titles based on topic similarity. These results are expected to assist in managing library collections and help users find relevant books more easily.

Keywords: text mining, clustering, K-Means, TF-IDF, Davies-Bouldin Index

DAFTAR ISI

KATA PENGANTAR.....	i
PERNYATAAN ORISINALITAS.....	ii
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS.....	iii
LEMBAR PENGESAHAN	iv
ABSTRAK	v
ABSTRACT	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	x
DAFTAR RUMUS.....	xi
DAFTAR TABEL	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Penelitian.....	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Luaran Yang Diharapkan	4
BAB II LANDASAN TEORI	5
2.1 <i>Data Mining</i>	5
2.1.1 Pengertian <i>Data Mining</i>	5
2.1.2 Kegunaan <i>Data Mining</i>	5
2.2 <i>Text Mining</i> dalam <i>Data Mining</i>	7
2.2.1 Perbedaan <i>Data Mining</i> dan <i>Text Mining</i>	7
2.2.2 Perbedaan <i>Data</i> Terstruktur dan Tidak Terstruktur.....	8
2.2.3 Kaitan <i>Text Mining</i> dalam <i>Data Mining</i>	9

2.3	Klustering (<i>Clustering</i>).....	10
2.4	Evaluasi <i>Davies-Bouldin Index</i>	11
2.5	Algoritma K-Means Clustering	11
2.6	<i>Deduplication</i>	12
2.7	<i>Text Pre-processing</i>	12
2.8	Penelitian Terdahulu	17
BAB III METODOLOGI PENELITIAN		20
3.1	Alur Penelitian	20
3.2	Tahapan Penelitian	21
3.3	Waktu dan Tempat Penelitian	23
3.4	Perangkat Keras dan Perangkat Lunak	23
3.4.1	Perangkat Keras	23
3.4.2	Perangkat Lunak	23
3.5	Jadwal Penelitian	23
BAB IV HASIL DAN PEMBAHASAN.....		25
4.1	<i>Data</i>	25
4.2.1	<i>Data Cleaning</i>	28
4.2.2	<i>Case Folding</i>	29
4.2.3	<i>Text Normalization</i>	30
4.2.4	<i>Deduplication</i>	32
4.2.5	<i>Text Cleanup / Punctual Removal</i>	33
4.2.6	<i>Tokenizing</i>	34
4.2.7	<i>Stopword Removal</i>	34
4.2.8	<i>Stemming</i>	38
4.2.9	<i>Second Data Cleaning</i>	40
4.3	<i>Featuring Engineering</i> dengan TF-IDF	41

4.4	Proses Klastering K-Means	47
4.5	Evaluasi.....	53
BAB V PENUTUP.....		62
5.1	Hasil dan Kesimpulan	62
5.2	Saran	62
DAFTAR PUSTAKA.....		64

DAFTAR GAMBAR

Gambar 2. 2 Contoh <i>Data</i> Terstruktur	6
Gambar 2. 3 Perbedaan <i>Data Mining</i> dan <i>Text Mining</i>	6
Gambar 3. 1 Alur Penelitian.....	26
Gambar 3. 2 Hasil <i>Dataset</i> yang didapat	27

DAFTAR RUMUS

Rumus <i>Davies-Bouldin Index</i>	11
Rumus <i>Euclidean Distance</i>	12
Rumus <i>Term Frequency</i>	40
Rumus <i>Inverse Document Frequency</i>	40
Rumus <i>Term Frequency - Inverse Document Frequency</i>	41
Rumus Jarak Dokumen D1 ke <i>Centroid C2</i>	49
Rumus <i>Elbow Method</i>	50

DAFTAR TABEL

Tabel 2.1 <i>Data terstruktur sebagai input ke metode supervised learning</i>	7
Tabel 2. 2 <i>Data terstruktur untuk input ke metode unsupervised learning</i>	8
Tabel 2. 3 Contoh <i>Token</i>	13
Tabel 2. 4 Contoh <i>Type</i>	13
Tabel 2. 5 Contoh <i>Term</i>	14
Tabel 2. 6 Tokenisasi sebuah kalimat Bahasa Indonesia.....	14
Tabel 2. 7 Hasil <i>Filtering</i>	15
Tabel 2. 8 Hasil <i>Stemming</i>	16
Tabel 2. 9 Hasil <i>Text Pre-processing</i> teks Bahasa Inggris	16
Tabel 2. 10 Penelitian Terkait <i>Klustering</i>	17
Tabel 3. 1 Jadwal Penelitian.....	23
Tabel 4. 1 <i>Data</i> peminjaman buku Universitas Pembangunan Nasional Veteran Jakarta	25
Tabel 4. 2 Hasil <i>Data Frame df_test</i>	26
Tabel 4. 3 <i>Data Cleaning</i>	27
Tabel 4. 4 Hasil <i>Case Folding</i>	28
Tabel 4. 5 Pratinjau <i>Dictionary</i> untuk <i>Text Normalization</i>	29
Tabel 4. 6 Perbedaan Dokumen dengan Kata Singkatan	30
Tabel 4. 7 Hasil <i>Text Normalization</i>	30
Tabel 4. 8 Perbandingan <i>Deduplication</i>	31
Tabel 4. 9 Hasil <i>Text Cleanup</i>	32
Tabel 4. 10 Hasil <i>Tokenizing</i>	33
Tabel 4. 11 <i>Stopword Removal</i> dengan Sastrawi.....	34
Tabel 4. 12 Pratinjau tambahan token_english.txt untuk data	34
Tabel 4. 13 Penerapan pratinjau token_english.txt terhadap <i>data</i>	35
Tabel 4. 14 Pratinjau <i>data</i> dengan <i>Stopword Addons</i>	36
Tabel 4. 15 Hasil Tahap Terakhir <i>Stopword Removal</i> dengan <i>Addons</i>	36

Tabel 4. 16 Pratinjau <i>Stemming Addons</i>	37
Tabel 4. 17 Hasil <i>Stemming</i> dengan <i>Dictionary</i>	38
Tabel 4. 18 Hasil <i>Stemming</i> dengan Sastrawi	38
Tabel 4. 19 Hasil <i>Cleaning Data yang Kedua</i>	39
Tabel 4. 20 Sampel Data Perhitungan TF-IDF	40
Tabel 4. 21 Perhitungan TF-IDF dengan lima sampel data.....	42
Tabel 4. 22 Parameter <i>Feature Engineering TfidfVectorizer</i>	45
Tabel 4. 23 Sampel <i>data</i> untuk Klastering	47
Tabel 4. 24 Kosaka <i>term</i>	47
Tabel 4. 25 Hasil vektor angka tiap dokumen.....	48
Tabel 4. 26 Hasil Perhitungan <i>Euclidean Distance</i> terhadap lima dokumen	49
Tabel 4. 27 <i>Parameter Elbow Method</i>	51
Tabel 4. 28 Hasil Visual Jarak <i>Centroid</i> dan <i>Cluster</i> terhadap lima dokumen	53
Tabel 4. 29 Hasil Nilai DBI untuk tiap <i>cluster</i>	55
Tabel 4. 30 Hasil Anggota dengan k=2	56
Tabel 4. 31 Hasil Anggota dengan k=3	56
Tabel 4. 32 Hasil Anggota dengan k=4	57
Tabel 4. 33 Hasil Anggota dengan k=5	57
Tabel 4. 34 Hasil Anggota dengan k=6	57
Tabel 4. 35 Hasil Anggota dengan k=7	58
Tabel 4. 36 Pratinjau Penamaan <i>Cluster</i>	59