

**KOMPARASI ALGORITMA K-MEANS DAN ALGORITMA
DBSCAN DALAM PREDIKSI BERITA HOAKS DI
INDONESIA**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Felicia Hanna Ratna

0000052746

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2024**

KOMPARASI ALGORITMA K-MEANS DAN ALGORITMA

DBSCAN DALAM PREDIKSI BERITA HOAKS DI

INDONESIA



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Diajukan Sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer (S.Kom)

Felicia Hanna Ratna

0000052746

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK DAN INFORMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2024

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Felicia Hanna Ratna

Nomor Induk Mahasiswa : 00000052746

Program studi : Sistem Informasi

Skripsi dengan judul:

“Komparasi Algoritma K-Means dan Algoritma DBSCAN dalam Prediksi Berita Hoaks di Indonesia”

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas Akhir yang telah saya tempuh.

Tangerang, 20 Desember 2023

UMMA



Felicia Hanna Ratna

UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PERSETUJUAN

Skripsi dengan judul

Komparasi Algoritma K-Means dan Algoritma DBSCAN dalam Prediksi Berita Hoaks di Indonesia

Oleh

Nama : Felicia Hanna Ratna
NIM : 00000052746
Program Studi : Sistem Informasi
Fakultas : Teknik & Informatika

Telah disetujui untuk diajukan pada
Sidang Ujian Skripsi Universitas Multimedia Nusantara

Tangerang, 13 Desember 2023

Pembimbing


Rudi Sutomo, S.Kom., M.Si.,
M.Kom. 0222057501

Ketua Program Studi Sistem Informasi


Ririn Ikana Desanti, S.Kom., M.Kom.

iii

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas academica Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

Nama : Felicia Hanna Ratna
NIM : 00000052746
Program Studi : Sistem Informasi
Fakultas : Teknik & Informatika
Jenis Karya : *Tesis/Skripsi/Tugas Akhir

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Multimedia Nusantara Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul.

**“KOMPARASI ALGORITMA K-MEANS DAN ALGORITMA DBSCAN
DALAM PREDIKSI BERITA HOAKS DI INDONESIA”**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini, Universitas Multimedia Nusantara berhak menyimpan, mengalihmediakan/mengalihformatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 13 Desember 2023

Yang menyatakan,



Felicia Hanna Ratna

KATA PENGANTAR

Puji Syukur atas selesainya penulisan skripsi ini dengan judul: “**KOMPARASI ALGORITMA K-MEANS DAN ALGORITMA DBSCAN DALAM PREDIKSI BERITA HOAKS DI INDONESIA**” dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar sarjana komputer Jurusan Sistem Informasi Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan tugas akhir ini, sangatlah sulit bagi saya untuk menyelesaikan tugas akhir ini. Oleh karena itu, saya mengucapkan terima kasih kepada

1. Dr. Ninok Leksono, M.A., selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Bapak Rudi Sutomo, S.Kom., M.Si., M.Kom., selaku pembimbing yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya laporan skripsi ini.
5. Teman-teman yang bersama-sama ikut serta dalam program *project independent acceleration* karena telah menjadi pendorong dan menjadi penyemangat dalam menyelesaikan laporan ini.
6. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Penulis menyadari bahwa penulisan laporan tugas akhir ini masih jauh dari kata sempurna sehingga penulis menerima segala kritik dan saran dari seluruh pembaca. Semoga karya ilmiah ini bermanfaat, baik sebagai sumber informasi maupun sebagai sumber inspirasi bagi para pembaca.

Tangerang, 13 Desember 2023



Felicia Hanna Ratna

KOMPARASI ALGORITMA K-MEANS DAN ALGORITMA

DBSCAN DALAM PREDIKSI BERITA HOAKS DI

INDONESIA

Felicia Hanna Ratna

ABSTRAK

Studi ini menggali peran platform X yang semakin meningkat dalam menyebarkan informasi sehingga muncul berita hoaks di lanskap digital. Kemudahan penyebaran konten yang tidak terverifikasi di platform X telah memicu kekhawatiran tentang penyebaran hoaks dan ujaran kebencian. Maka dari itu, dilakukan penelitian ini dengan tujuan untuk mengevaluasi perbandingan algoritma K-Means dan DBSCAN menggunakan metode penelitian, yaitu KDD dengan alur penelitian yang berisikan tahap data selection, pre-processing, transformation, data mining, dan evaluation. Hal ini dilakukan untuk melihat dan memahami pola penyebaran berita hoaks di Indonesia. Hasil penelitian ini menunjukkan bahwa meskipun K-Means menghasilkan cluster yang lebih erat dengan performa nilai Davies-Bouldin Index sebesar 0,38, namun gagal dalam memisahkan cluster dengan nilai silhouette score sebesar 0,72 dan SSE sebesar 47,708. Sebaliknya, DBSCAN menghasilkan pemisahan kluster yang lebih jelas dengan nilai DBI yang lebih tinggi, yakni mencapai 7,06. Sehingga dapat disimpulkan, algoritma K-Means dan DBSCAN miliki keunggulan dan kelemahan unik dalam klustering data. K-Means menghasilkan kluster yang jelas dengan fokus pada pusat kluster, namun sulit memisahkan kluster yang tumpang tindih. Sementara DBSCAN tegas dalam pemisahan berdasarkan kepadatan, namun kurang seragam dalam kluster yang dihasilkan, lebih cocok untuk data tak teratur.

Kata kunci: Berita Hoaks, DBSCAN, K-Means, *Knowledge Discovery in Database*

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

COMPARISON OF K-MEANS ALGORITHM AND DBSCAN

ALGORITHM IN PREDICTING OF HOAX NEWS IN

INDONESIA

Felicia Hanna Ratna

ABSTRACT (English)

This study explores the increasing role of platform X in disseminating information, resulting in the emergence of hoax news in the digital landscape. The ease of spreading unverified content on platform X has sparked concerns about the spread of hoaxes and hate speech. Therefore, this research was conducted with the aim of evaluating the comparison of the K-Means and DBSCAN algorithms using a research method, namely KDD with a research flow containing the stages of data selection, pre-processing, transformation, data mining, and evaluation. This is done to see and understand the pattern of hoax news distribution in Indonesia. The results of this study show that although K-Means produces tighter clusters with a Davies-Bouldin Index score performance of 0.38, it fails to separate clusters with a silhouette score of 0.72 and SSE of 47,708. In contrast, DBSCAN produces a clearer cluster separation with a higher DBI value, which reaches 7.06. So, it can be concluded, the K-Means and DBSCAN algorithms have unique advantages and disadvantages in clustering data. K-Means produces clear clusters by focusing on the cluster center, but it is difficult to separate overlapping clusters. While DBSCAN is firm in density-based separation, it is less uniform in the clusters produced, more suitable for irregular data.

Keywords: Hoax News, DBSCAN, K-Means, Knowledge Discovery in Database

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	iv
KATA PENGANTAR.....	v
ABSTRAK	vi
<i>ABSTRACT (English)</i>	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah.....	5
1.4 Tujuan dan Manfaat Penelitian	6
1.4.1 Tujuan Penelitian.....	6
1.4.2 Manfaat Penelitian.....	6
1.5 Sistematika Penulisan	7
BAB II LANDASAN TEORI	8
2.1 Algoritma dan Framework yang digunakan	23
2.2 Tools dan Software yang digunakan	24
2.3 Penelitian Terdahulu.....	29
BAB III METODOLOGI PENELITIAN	32
3.1 Gambaran Umum Objek Penelitian.....	32
3.2 Metode Penelitian	32
3.2.1 Alur Penelitian.....	33
3.3 Teknik Pengumpulan Data.....	39
3.3.1 Populasi dan Sampel.....	39
3.3.2 Periode Pengambilan Data	40

3.4	Teknik Analisis Data	40
BAB IV ANALISIS DAN HASIL PENELITIAN		43
4.1	Data Selection	47
4.2	Pre-processing.....	51
4.2.1	Import Library	51
4.2.2	Transform into Dataframe	52
4.2.3	Data Cleansing.....	54
4.3	Transformation.....	57
4.4	Exploratory Data Analysis (EDA)	58
4.5	Social Network Analysis (SNA)	60
4.6	Data Mining	65
4.7	Evaluation	69
4.8	Hasil dan Diskusi	71
BAB V SIMPULAN DAN SARAN		73
5.1	Simpulan.....	73
5.2	Saran.....	74
DAFTAR PUSTAKA		76
LAMPIRAN.....		83

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	29
Tabel 3.1 Perbandingan Metode Data Mining	32
Tabel 3.2 Variabel Sheet Edges	35
Tabel 3.3 Variabel Sheet Vertices.....	36
Tabel 3.4 Perbandingan Google Colab dan Jupyter Notebook.....	39
Tabel 3.5 Teknik Analisis Data.....	40
Tabel 4.1 Keterangan Profil Pengguna	43
Tabel 4.2 Social Network Analysis.....	64
Tabel 4.3 Tabel evaluasi performa.....	70

A large, light blue circular watermark logo is centered on the page. It features a stylized 'U' shape on the left and a grid of squares on the right, resembling a data visualization or a network diagram.

UMMN

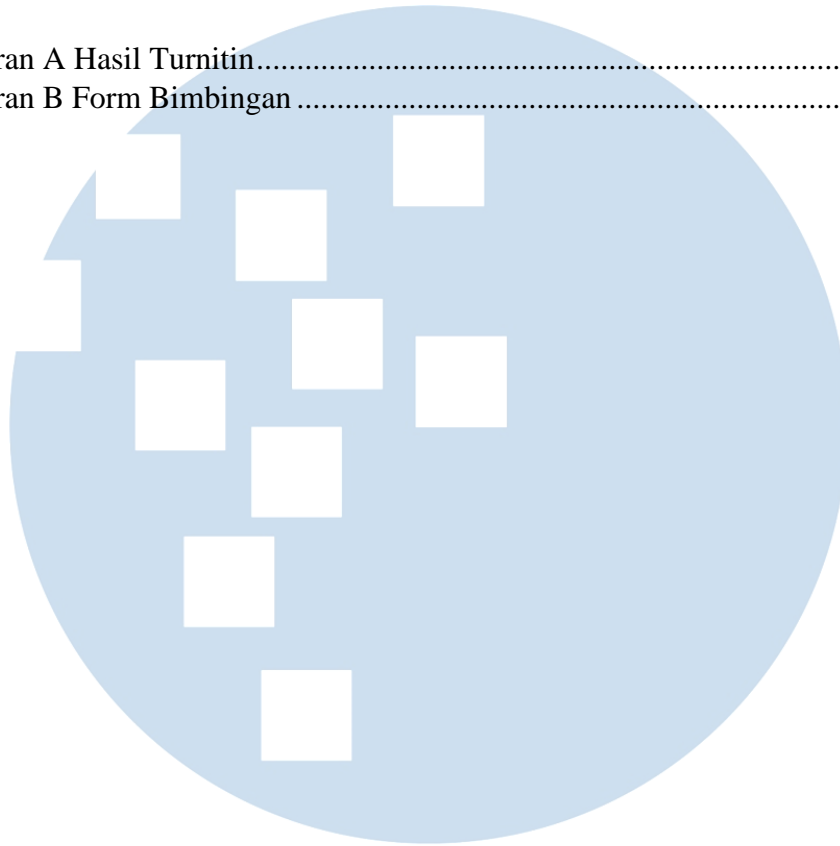
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR GAMBAR

Gambar 2.1 Tahap Knowledge Discovery in Databases Sumber: Javatpoint.....	23
Gambar 2.2 Logo Python Sumber: PNGEgg	25
Gambar 2.3 Logo Google Colab Sumber: wikipedia	26
Gambar 2.4 Logo Microsoft Excel Sumber: Logo.wine	27
Gambar 2.5 Logo X Sumber: twitter.com	27
Gambar 2.6 Logo NodeXL Sumber: nodexl.com	28
Gambar 3.1 Alur Penelitian.....	34
Gambar 4.1 Contoh Profil Pengguna X	43
Gambar 4.2 Contoh Relasi Pengguna X	45
Gambar 4.3 Contoh Data Jaringan	46
Gambar 4.4 Tabel Data Sheet Edges.....	49
Gambar 4.5 Tabel Data Sheet Vertices	50
Gambar 4.6 Import Library	52
Gambar 4.7 Code untuk membaca data sheet edges	53
Gambar 4.8 Code untuk membaca data sheet vertices.....	53
Gambar 4.9 Membuat dataframe di sheet edges	53
Gambar 4.10 Membuat dataframe di sheet vertices.....	53
Gambar 4.11 Mengecek missing value sheet edges.....	54
Gambar 4.12 Mengecek missing value sheet vertices	54
Gambar 4.13 Menghapus kolom sheet edges.....	56
Gambar 4.14 Menghapus kolom sheet vertices	56
Gambar 4.15 Mengkonversi kolom relationship date	56
Gambar 4.16 Menghapus kolom sheet edges.....	57
Gambar 4.17 Menghapus kolom sheet vertices	57
Gambar 4.18 Mengubah data vertex	57
Gambar 4.19 Reduksi dimensi dengan PCA.....	58
Gambar 4.20 Plot Jumlah Tweet per tanggal	59
Gambar 4.21 Peringkat pengguna dengan tweet terbanyak.....	60
Gambar 4.22 Degree Centrality	60
Gambar 4.23 Betweenness centrality	61
Gambar 4.24 Closeness centrality	62
Gambar 4.25 Eigenvector centrality	63
Gambar 4.26 Contoh akun X yang dinyatakan sebagai suspicious hoax dan non suspicious hoax	64
Gambar 4.27 Plot elbow method.....	66
Gambar 4.28 Clustering K-Means dan DBSCAN tanpa PCA.....	67
Gambar 4.29 Clustering K-Means dan DBSCAN dengan PCA	68

DAFTAR LAMPIRAN

Lampiran A Hasil Turnitin.....	83
Lampiran B Form Bimbingan	93



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA