

**SISTEM REKOMENDASI IDE KONTEN OTOMATIS UNTUK *CONTENT*
CREATOR BERBASIS TREN DAN *ENGAGEMENT* KONTEN
MENGUNAKAN *TF-IDF* DAN *COSINE SIMILARITY***

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Sistem Informasi



disusun oleh

KAMILATURRAHMAH MD

22.12.2584

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**SISTEM REKOMENDASI IDE KONTEN OTOMATIS UNTUK *CONTENT*
CREATOR BERBASIS TREN DAN *ENGAGEMENT* KONTEN
MENGUNAKAN *TF-IDF* DAN *COSINE SIMILARITY***

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

KAMILATURRAHMAH MD

22.12.2584

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**SISTEM REKOMENDASI IDE KONTEN OTOMATIS UNTUK *CONTENT*
CREATOR BERBASIS TREN DAN *ENGAGEMENT* KONTEN
MENGUNAKAN *TF-IDF* DAN *COSINE SIMILARITY***

yang disusun dan diajukan oleh

KAMILATURRAHMAH MD

22.12.2584

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 6 Maret 2026

Dosen Pembimbing,



Ir. Yoga Pristyanto, S.Kom., M.Eng
NIK. 190302412

HALAMAN PENGESAHAN

SKRIPSI

**SISTEM REKOMENDASI IDE KONTEN OTOMATIS UNTUK *CONTENT*
CREATOR BERBASIS TREN DAN *ENGAGEMENT* KONTEN
MENGUNAKAN *TF-IDF* DAN *COSINE SIMILARITY***

yang disusun dan diajukan oleh

KAMILATURRAHMAH MD

22.12.2584

Telah dipertahankan di depan Dewan Penguji
pada tanggal 6 Maret 2026

Susunan Dewan Penguji

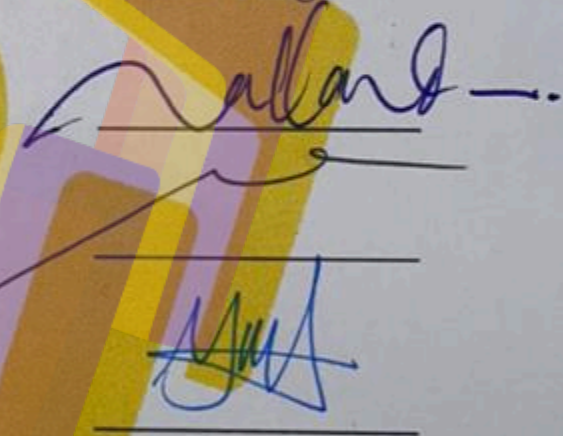
Nama Penguji

Tanda Tangan

Akhmad Dahlan, S.Kom., M.Kom.
NIK. 190302174

Ibnu Hadi Purwanto, S.Kom., M.Kom.
NIK. 190302390

Ir. Yoga Pristyanto, S.Kom., M.Eng.
NIK. 190302412



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 26 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Kamilaturrahmah MD
NIM : 22.12.2584

Menyatakan bahwa Skripsi dengan judul berikut:

Sistem rekomendasi ide konten otomatis untuk *content creator* berbasis tren dan *engagement* konten menggunakan *tf-idf* dan *cosine similarity*

Dosen Pembimbing : Ir. Yoga Pristyanto, S.Kom., M.Eng.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 26 Februari 2026

Yang Menyatakan,


Kamilaturrahmah MD

HALAMAN PERSEMBAHAN

Dengan penuh kerendahan hati dan rasa syukur kepada Allah SWT atas segala rahmat, kekuatan, dan kesabaran yang senantiasa mengiringi setiap langkah dalam perjalanan ini.

Saya persembahkan karya ini kepada kedua orang tua tercinta, yang tidak pernah berhenti mendoakan, mendukung, dan menguatkan saya dalam setiap keadaan. Terima kasih atas kasih sayang yang tulus, pengorbanan yang tak terbilang jumlahnya, serta kepercayaan yang selalu diberikan, hingga saya mampu sampai pada titik ini. Kepada Bapak/Ibu Dosen wali, pengampu mata kuliah dan tentunya Bapak Yoga Pristyanto, S.Kom., M.Eng selaku pembimbing, terima kasih atas arahan, pengertian dan bimbingan yang telah diberikan. Setiap nasihat dan pembelajaran menjadi bekal berharga dalam proses saya bertumbuh, baik secara akademik maupun pribadi. Kepada teman-teman seperjuangan, terima kasih atas kebersamaan, semangat, dan dukungan yang terselip di setiap proses ini. Kehadiran kalian menjadi penghibur di saat jenuh, sekaligus pengingat bahwa perjuangan ini tidak pernah dijalani sendirian.

Dan kepada diri saya sendiri, untuk setiap keraguan yang berhasil dilampaui, untuk setiap lelah yang tetap dipaksa melangkah, serta ego yang pernah goyah namun terus belajar bertahan, karya ini menjadi saksi bahwa proses selama 4 tahun ini bukan sia-sia.

Akhirnya, skripsi ini saya persembahkan dengan sepenuh hati, terutama untuk orang tua saya, sebagai bentuk bakti, terima kasih, dan harapan kecil agar apa yang telah saya capai hari ini dapat menjadi kebanggaan dan kebahagiaan bagi mereka. Terima kasih

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah Swt. atas segala rahmat dan karunia-Nya sehingga skripsi ini dapat diselesaikan sebagai salah satu syarat untuk memperoleh gelar sarjana.

Penyusunan skripsi ini tidak terlepas dari bimbingan dan dukungan berbagai pihak. Penulis menyampaikan terima kasih kepada Bapak/Ibu Yoga Pristyanto, S.Kom., M.Eng selaku dosen pembimbing atas arahan, kesabaran, dan ilmu yang diberikan selama proses penyusunan skripsi. Terima kasih juga kepada seluruh dosen dan staf akademik Universitas Amikom Yogyakarta atas ilmu dan dukungan selama masa perkuliahan.

Dengan penuh rasa hormat, penulis juga menyampaikan terima kasih kepada kedua orang tua tercinta atas doa, kasih sayang, dan dukungan yang tidak pernah terputus hingga penulis mampu menyelesaikan tahap ini.

Penulis menyadari bahwa proses penyusunan skripsi ini merupakan perjalanan yang tidak mudah. Banyak pembelajaran, tantangan, serta proses pendewasaan yang dilalui.

Semoga segala bantuan, dukungan, dan kebaikan yang telah diberikan mendapatkan balasan yang berlipat dari Allah Swt. Penulis berharap skripsi ini dapat memberikan manfaat serta menjadi bagian kecil dari kontribusi dalam pengembangan ilmu pengetahuan.

Yogyakarta, 10 Februari 2026

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN.....	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xv
INTISARI	xvii
<i>ABSTRACT</i>	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	12
2.2.1 <i>Data Mining</i>	12
2.2.2 <i>Intelligent Business</i>	13
2.2.3 Sistem Rekomendasi.....	14
2.2.4 Media Sosial dan Tren Konten	15
2.2.5 <i>Engagement Konten</i>	16
2.2.6 <i>Natural Language Processing (NLP)</i>	16

2.2.7	<i>Term Frequency–Inverse Document Frequency (TF-IDF)</i>	17
2.2.8	<i>Cosine Similarity</i>	18
2.2.9	<i>Website</i> sebagai Media Implementasi Sistem	19
BAB III METODE PENELITIAN		21
3.1	Objek Penelitian	21
3.2	Alur Penelitian	21
3.2.1	Akuisisi Data	24
3.2.2	<i>Preprocessing</i> Data	25
3.2.3	Penentuan Konten Populer	28
3.2.4	Representasi Teks Menggunakan <i>TF-IDF</i>	32
3.2.5	Pengukuran Kemiripan Menggunakan <i>Cosine Similarity</i>	32
3.2.6	Pembuatan Model Rekomendasi	33
3.2.7	Implementasi Sistem Berbasis <i>Web (Deployment)</i>	33
3.2.8	Pengujian Fungsional Sistem	34
3.2.9	Evaluasi dan Analisis Hasil	34
SISTEM REKOMENDASI		35
3.2.10	Akses sistem	35
3.2.11	<i>Login</i> dan Registrasi Pengguna	35
3.2.12	Pemilihan <i>Platform</i> Konten	35
3.2.13	<i>Input</i> Kata Kunci atau Ide Konten	35
3.2.14	<i>Preprocessing</i> Teks	35
3.2.15	Representasi Teks Menggunakan <i>TF-IDF</i>	35
3.2.16	Pengukuran Kemiripan Menggunakan <i>Cosine Similarity</i>	36
3.2.17	Pembuatan Model Rekomendasi	36
3.2.18	Penyajian Rekomendasi Ide Konten	36
3.2.19	Penyimpanan Rekomendasi ke Favorit	36
3.2.20	<i>Logout</i> dan Penyelesaian	36
3.3	Alat dan Bahan	36
3.3.1	Data Penelitian	37
3.3.2	Alat dan Instrumen Penelitian	37
BAB IV HASIL DAN PEMBAHASAN		39
4.1	Akuisisi dan Pra-pemrosesan Data	39

4.1.1 Hasil Implementasi	44
4.2 Implementasi Sistem <i>website</i>	45
4.2.1 Antarmuka Pengguna.....	45
4.2.2 Fitur Pencarian dan Rekomendasi	48
4.2.3 Fitur Favorit	51
4.2.4 <i>Logout</i> dan <i>Session Management</i>	53
4.2.5 <i>Responsive Design (Mobile View)</i>	54
4.2.6 <i>Deployment</i> dan <i>Production Environment</i>	56
4.3.1 Pengujian Fungsional (<i>Black Box Testing</i>).....	58
4.3.2 Pengujian Performa Sistem.....	66
4.3.3 Analisis Statistik Hasil Pengujian.....	68
4.3.4 Pengujian <i>Usability</i> dengan <i>System Usability Scale (SUS)</i>	72
4.4 Pembahasan	85
4.4.1 Analisis Hasil Rekomendasi	85
4.4.2 Evaluasi Sistem.....	85
4.4.3 Perbandingan dengan Penelitian Terdahulu	88
BAB V PENUTUP	91
5.1 Kesimpulan	91
5.2 Saran	92
REFERENSI	94
LAMPIRAN.....	96

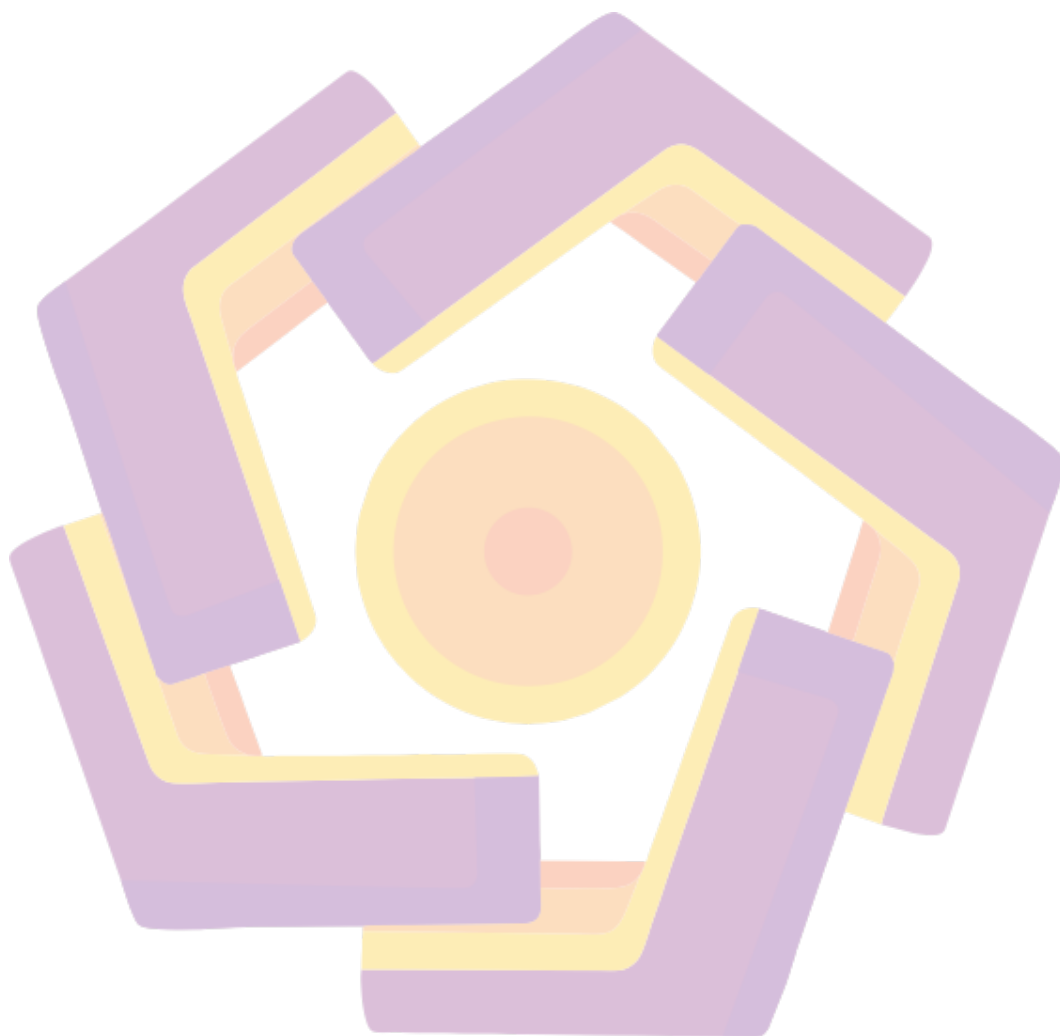
DAFTAR TABEL

Tabel 2.1. Keaslian Penelitian	8
Tabel 4.1. Skenario Pengujian Fitur Autentikasi	58
Tabel 4.2. Skenario Pengujian Fitur Pencarian dan Rekomendasi	60
Tabel 4.3. Skenario Pengujian Fitur Favorit	63
Tabel 4.4. Skenario Pengujian UI/UX	64
Tabel 4.5. Pengujian <i>Response Time API Endpoints</i>	65
Tabel 4.6. Pengujian Akurasi dan Relevansi Rekomendasi	66
Tabel 4.7. Ringkasan Statistik Hasil Pengujian <i>Black Box</i>	67
Tabel 4.8. Analisis <i>Precision</i> Rekomendasi	68
Tabel 4.9. Statistik <i>Response Time</i> (detik)	69
Tabel 4.10. Analisis <i>Error Rate</i> dan <i>Reliability</i>	70
Tabel 4.11. Profil Status Responden Pengujian SUS	72
Tabel 4.12. Instrumen Kuesioner SUS <i>Content Finder</i>	73
Tabel 4.13. Rekap Skor SUS Seluruh Responden	76
Tabel 4.14. Statistik Deskriptif Skor SUS	79
Tabel 4.15. Distribusi <i>Grade Usability Sistem Content Finder</i>	80
Tabel 4.16. Rata-rata Skor per Pertanyaan SUS	81
Tabel 4.17. Perbandingan Skor SUS dengan Benchmark Industri	83
Tabel 4.18. Perbandingan dengan Penelitian Terdahulu	87

DAFTAR GAMBAR

Gambar 3.1. <i>Flowchart</i> pengolahan data	21
Gambar 3.2. <i>Flowchart</i> sistem rekomendasi	22
Gambar 3.3. Pengambilan data dengan YouTube Data API v3	24
Gambar 3.4. Pengambilan data dengan <i>scrapping</i> menggunakan layanan Apify	24
Gambar 4.1. Isi dataset	38
Gambar 4.2. Isi kolom text (hasil penggabungan judul dan hashtag)	39
Gambar 4.3. Sesudah dan sebelum noise removal	39
Gambar 4.4. Hasil segmentasi kata	40
Gambar 4.5. Hasil tokenisasi	40
Gambar 4.6. Hasil perhitungan <i>engagement</i> Youtube	41
Gambar 4.7. Hasil penentuan <i>threshold</i>	41
Gambar 4.8. Hasil perhitungan <i>watchtime</i>	41
Gambar 4.9. Hasil penentuan konten populer	42
Gambar 4.10. Contoh matriks TF-IDF	42
Gambar 4.11. Contoh kemiripan konten pertama dengan 10 konten lainnya	42
Gambar 4.12. Sintak kode <i>hybrid ranking</i>	43
Gambar 4.13. Hasil implementasi model sistem rekomendasi konten populer	43
Gambar 4.14. Halaman Awal (<i>Landing Page</i>)	44
Gambar 4.15. Form Registrasi Pengguna	45
Gambar 4.16. Form <i>Login</i> dengan Fitur "Ingat Saya"	46
Gambar 4.17. <i>Interface</i> Pencarian Konten	47
Gambar 4.18. <i>Loading State</i> saat Pemrosesan	48
Gambar 4.19. Hasil Rekomendasi Konten	49
Gambar 4.20. <i>Empty State</i> - Tidak Ada Hasil	49
Gambar 4.21. Tab Favorit (Kosong)	50
Gambar 4.22. Tab Favorit (Berisi Konten)	51
Gambar 4.23. <i>Alert</i> - Ditambahkan ke Favorit	51
Gambar 4.24. <i>Confirm Dialog</i> - Hapus dari Favorit	52
Gambar 4.25. <i>Confirm Dialog</i> – Logout	52
Gambar 4.26. <i>Mobile View</i> - <i>Landing Page</i>	53
Gambar 4.27. <i>Mobile View</i> - Pencarian dan Hasil	54
Gambar 4.28. CloudHebat cPanel - <i>Python App Management</i>	55
Gambar 4.29. <i>Database Management</i> via <i>phpMyAdmin</i>	56

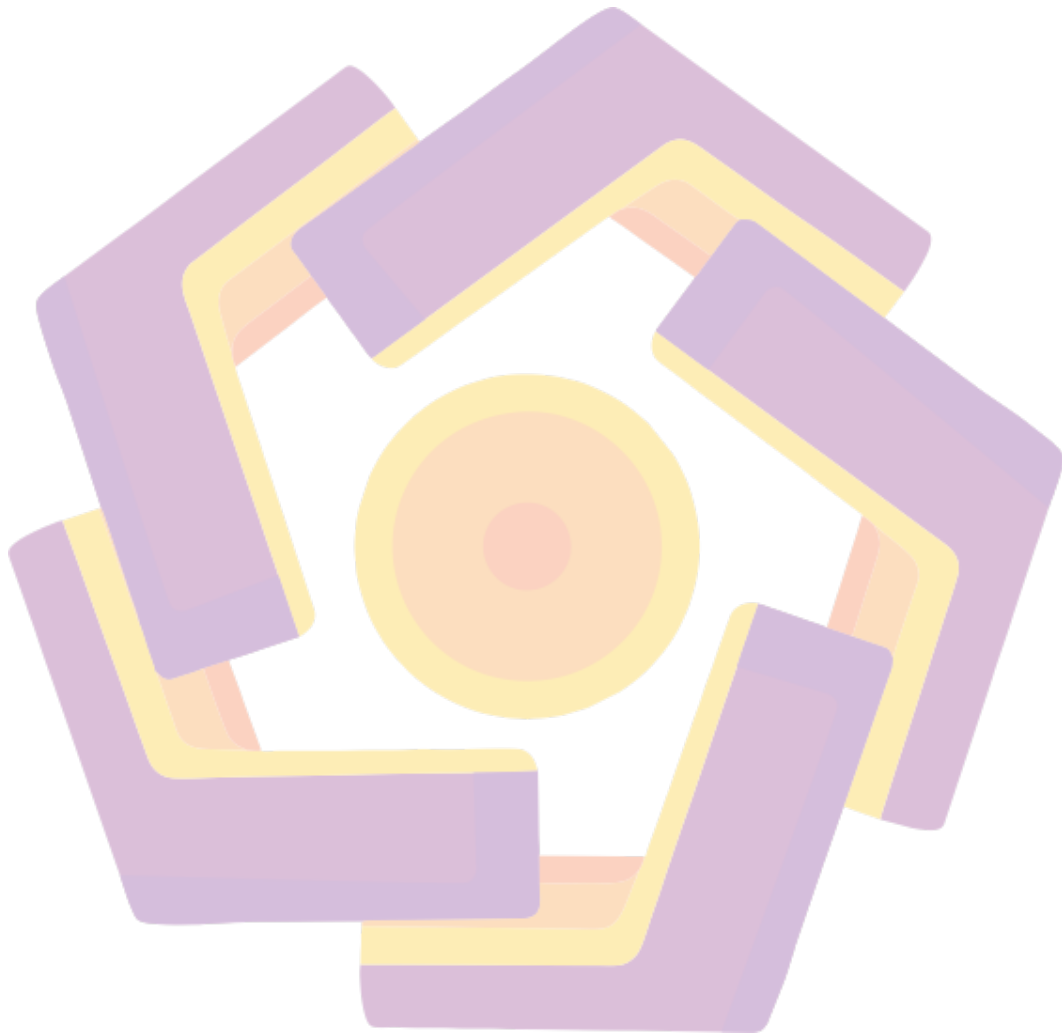
DAFTAR LAMPIRAN



DAFTAR LAMBANG DAN SINGKATAN

Σ	Sigma (simbol penjumlahan dalam matematika dan statistik)
μ	Mu (simbol mean/rata-rata dalam statistik)
$\times$	Simbol perkalian
$\geq$	Lebih besar atau sama dengan
$\leq$	Lebih kecil atau sama dengan
<i>API</i>	<i>Application Programming Interface</i>
<i>CPU</i>	<i>Central Processing Unit</i>
<i>CSV</i>	<i>Comma-Separated Values</i>
<i>HTML</i>	<i>HyperText Markup Language</i>
<i>HTTP</i>	<i>HyperText Transfer Protocol</i>
<i>HTTPS</i>	<i>HyperText Transfer Protocol Secure</i>
<i>IT</i>	<i>Information Technology</i>
<i>JWT</i>	<i>JSON Web Token</i>
<i>KDD</i>	<i>Knowledge Discovery in Databases</i>
<i>ML</i>	<i>Machine Learning</i>
<i>MySQL</i>	<i>My Structured Query Language</i>
<i>NLP</i>	<i>Natural Language Processing</i>
<i>RAM</i>	<i>Random Access Memory</i>
<i>REST</i>	<i>Representational State Transfer</i>
<i>SDK</i>	<i>Software Development Kit</i>
<i>SQL</i>	<i>Structured Query Language</i>
<i>SSL</i>	<i>Secure Sockets Layer</i>
<i>SUS</i>	<i>System Usability Scale</i>
<i>SVM</i>	<i>Support Vector Machines</i>
<i>TF-IDF</i>	<i>Term Frequency-Inverse Document Frequency</i>
<i>UI</i>	<i>User Interface</i>
<i>URL</i>	<i>Uniform Resource Locator</i>

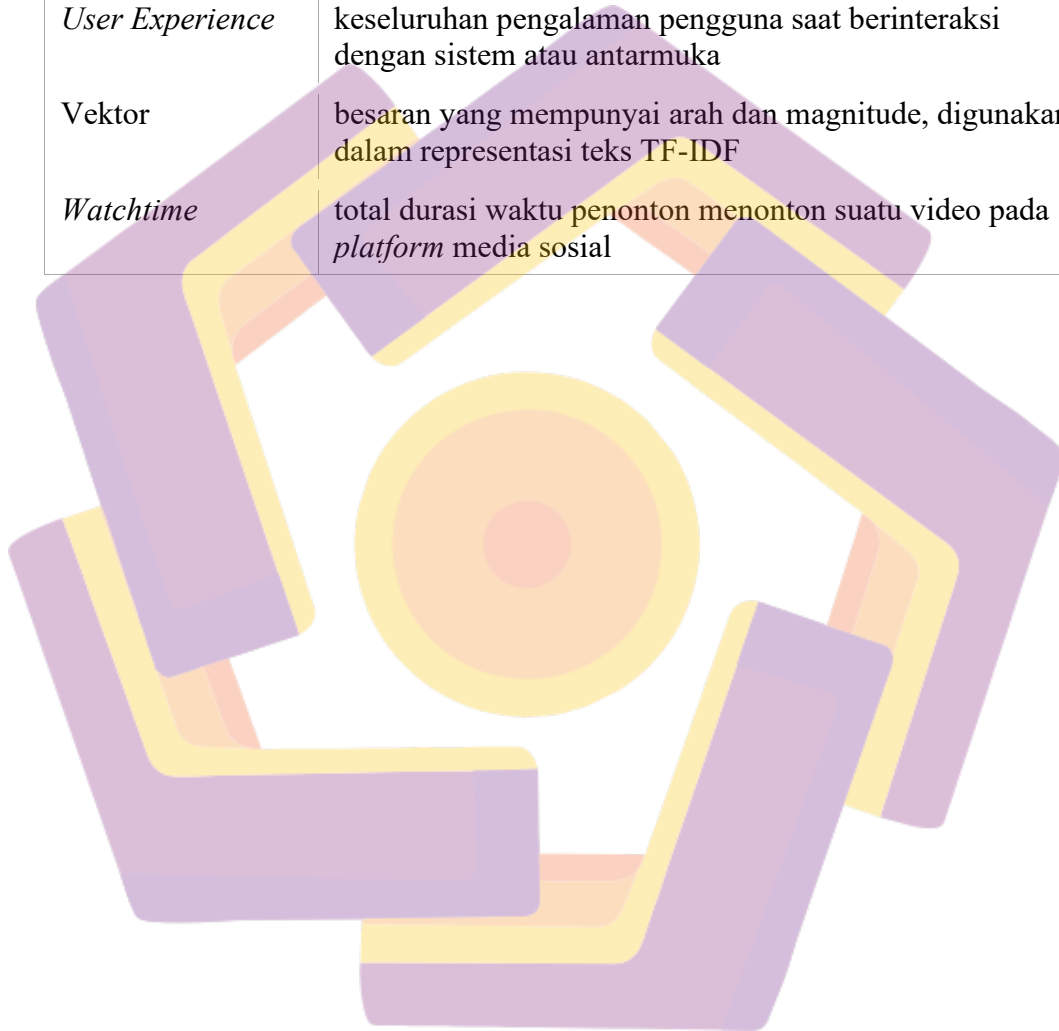
<i>UX</i>	<i>User Experience</i>
<i>WSGI</i>	<i>Web Server Gateway Interface</i>



DAFTAR ISTILAH

<i>Backend</i>	bagian <i>server-side</i> dari aplikasi web yang mengelola logika bisnis dan pengelolaan data
<i>Benchmarking</i>	proses membandingkan kinerja sistem dengan standar industri atau sistem pembanding lainnya
<i>Deployment</i>	proses menempatkan aplikasi ke <i>server</i> produksi sehingga dapat diakses oleh publik
<i>Eigen Value</i>	akar-akar persamaan karakteristik dari suatu matriks yang digunakan dalam analisis linear
<i>Engagement Rate</i>	rasio interaksi pengguna terhadap konten dibandingkan jumlah tayangan atau pengikut
<i>Frontend</i>	bagian <i>client-side</i> dari aplikasi web yang berinteraksi langsung dengan pengguna
<i>Grade</i>	tingkat penilaian atau kategori kualitas sistem berdasarkan rentang skor tertentu
<i>Interpretasi</i>	proses memahami dan menjelaskan makna dari data atau hasil pengujian yang diperoleh
<i>Likert Scale</i>	skala pengukuran sikap yang menggunakan beberapa tingkat persetujuan (1 hingga 5)
<i>Precision</i>	proporsi hasil rekomendasi yang relevan dari total hasil yang diberikan oleh sistem
<i>Preprocessing</i>	tahap persiapan dan pembersihan data sebelum diproses lebih lanjut oleh algoritma
Responden	individu yang memberikan tanggapan atau jawaban dalam survei atau pengujian <i>usability</i>
<i>Response Time</i>	waktu yang dibutuhkan sistem untuk merespons permintaan (<i>request</i>) dari pengguna
<i>Standard Deviation</i>	ukuran sebaran atau variasi data dari nilai rata-rata dalam analisis statistik
<i>Stemming</i>	proses mengubah kata berimbuhan menjadi kata dasar menggunakan algoritma linguistik
<i>SUS Score</i>	nilai numerik antara 0-100 yang menunjukkan tingkat <i>usability</i> suatu sistem

<i>Threshold</i>	nilai batas yang digunakan untuk <i>filtering</i> atau klasifikasi dalam pengolahan data
Tokenisasi	proses memecah teks menjadi unit-unit kecil (token) berupa kata atau frasa
<i>Usability</i>	tingkat kemudahan dan efektivitas sistem dalam membantu pengguna mencapai tujuan
<i>User Experience</i>	keseluruhan pengalaman pengguna saat berinteraksi dengan sistem atau antarmuka
Vektor	besaran yang mempunyai arah dan magnitude, digunakan dalam representasi teks TF-IDF
<i>Watchtime</i>	total durasi waktu penonton menonton suatu video pada platform media sosial



INTISARI

Perkembangan pesat media sosial mendorong meningkatnya jumlah *content creator* yang dituntut untuk menghasilkan konten secara konsisten, relevan, dan sesuai dengan tren yang sedang berkembang. Namun, proses pencarian ide konten masih banyak dilakukan secara manual dengan mengamati konten populer, sehingga memerlukan waktu yang cukup lama dan kurang efisien. Permasalahan tersebut menunjukkan perlunya sebuah sistem yang mampu membantu *content creator* dalam menghasilkan ide konten secara otomatis dan berbasis data.

Penelitian ini bertujuan untuk membangun sistem rekomendasi ide konten otomatis berbasis *website* dengan memanfaatkan konsep *Data Mining* dan *Intelligent Business*. Sistem ini menggunakan data konten populer dari *platform* media sosial dengan indikator *engagement* berupa jumlah penayangan (*views*), suka (*likes*), bagikan (*shares*), simpan (*saves*), jumlah waktu tayang (*watchtime*). Data teks berupa *caption* dan *hashtag* diolah menggunakan teknik *Natural Language Processing (NLP)* melalui tahapan *preprocessing* teks. Selanjutnya, representasi teks dilakukan menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF)*, sedangkan proses rekomendasi ditentukan berdasarkan tingkat kemiripan menggunakan metode *Cosine Similarity*.

Hasil dari penelitian ini berupa sistem yang mampu memberikan rekomendasi ide konten yang relevan, lengkap dengan tema, *caption*, dan *hashtag*, serta menyediakan fitur penyimpanan atau favorit rekomendasi. Sistem ini diharapkan dapat menjadi alat bantu pengambilan keputusan bagi *content creator* dalam menentukan ide konten yang sesuai dengan tren dan memiliki potensi *engagement* tinggi.

Kata kunci: *data mining*, sistem rekomendasi, ide konten, *TF-IDF*, *cosine similarity*.

ABSTRACT

The API growth of social media has increased the number of content creators who are required to consistently produce relevant and trend-aligned content. However, the process of generating content ideas is often conducted manually by observing popular content, which is time-consuming and inefficient. This condition indicates the need for a system that can automatically generate data-driven content ideas.

This study aims to develop a web-based automatic content idea recommendation system by applying the concepts of Data Mining and Intelligent Business. The system utilizes popular content data from social media platforms with engagement indicators such as views, likes, shares, saves and watchtimes. Textual data in the form of captions and hashtags are processed using Natural Language Processing (NLP) techniques through text preprocessing stages. The text representation is performed using the Term Frequency–Inverse Document Frequency (TF-IDF) method, while the recommendation process is based on similarity measurement using the Cosine Similarity method.

The results of this study produce a system capable of providing relevant content idea recommendations, including themes, captions, and hashtags, as well as a save or favorite feature. This system is expected to support content creators as a decision-support tool in selecting content ideas that align with trends and have high engagement potential.

Keyword: data mining, recommendation system, content ideas, TF-IDF, cosine similarity.