

**SISTEM REKOMENDASI CONTENT-BASED FILTERING
MENGUNAKAN SBERT DAN COSINE SIMILARITY UNTUK
REKOMENDASI ARTIKEL BERITA**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

ONNY VERGIEAN SAPUTRA

19.11.2731

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

**SISTEM REKOMENDASI CONTENT-BASED FILTERING
MENGUNAKAN SBERT DAN COSINE SIMILARITY UNTUK
REKOMENDASI ARTIKEL BERITA**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

ONNY VERGIEAN SAPUTRA

19.11.2731

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

HALAMAN PERSETUJUAN

SKRIPSI

**SISTEM REKOMENDASI CONTENT-BASED FILTERING MENGGUNAKAN
SBERT DAN COSINE SIMILARITY UNTUK REKOMENDASI ARTIKEL
BERITA**

yang disusun dan diajukan oleh

Onny Vergiean Saputra

19.11.2731

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 04 Agustus 2023

Dosen Pembimbing,



Arif Akbarul Huda, S.Si, M.Eng

NIK. 190302287

HALAMAN PENGESAHAN

SKRIPSI

**SISTEM REKOMENDASI CONTENT-BASED FILTERING MENGGUNAKAN
SBERT DAN COSINE SIMILARITY UNTUK REKOMENDASI ARTIKEL**

BERITA

yang disusun dan diajukan oleh

Onny Vergiean Saputra

19.11.2731

Telah dipertahankan di depan Dewan Penguji
pada tanggal 04 Agustus 2023

Susunan Dewan Penguji

Nama Penguji

Dr. Ferry Wahyu Wibowo, S.Si., M.Cs.
NIK. 190302235

Agit Amrullah, S.Kom., M.Kom.
NIK. 190302356

Arif Akbarul Huda, S.Si., M.Eng.
NIK. 190302287

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 04 Agustus 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Onny Vergiean Saputra
NIM : 19.11.2731

Menyatakan bahwa Skripsi dengan judul berikut:

Sistem Rekomendasi Content-based Filtering Menggunakan SBERT dan Cosine Similarity untuk Rekomendasi Artikel Berita

Dosen Pembimbing: Arif Akbarul Huda, S.Si, 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, 04 Agustus 2023

Yang Menyatakan,



The image shows an official stamp of AMIKOM Yogyakarta. The stamp includes the university's logo, the name 'AMIKOM YOGYAKARTA', and the acronym 'AMIKOM'. Below the stamp, there is a handwritten signature in black ink.

Onny Vergiean Saputra

KATA PENGANTAR

Dengan memanjatkan puja dan puji syukur kehadiran Allah Subhanahu Wata'ala, yang telah memberikan nikmat ilmu, kesehatan dan kecukupan rezeki sehingga penulis mampu dalam menyelesaikan skripsi ini yang berjudul "Sistem Rekomendasi Content-based Filtering Menggunakan SBERT dan Cosine Similarity untuk Rekomendasi Artikel Berita", sebagai salah satu syarat dalam menyelesaikan Program Sarjana (SI) Jurusan Informatika.

Penulis sadar bahwa jika tanpa bantuan dan dukungan dari berbagai pihak penyusunan skripsi ini tidak akan berjalan lancar, oleh karena itu penulis ingin menyampaikan rasa terima kasih sebesar-besarnya kepada seluruh pihak terkait yang telah memberikan dukungan dan motivasi dalam penyusunan skripsi ini. Pada kesempatan ini penulis ingin menyampaikan terima kasih kepada:

1. Kedua orang tua tercinta yang selalu memberikan dukungan penuh, baik secara materi maupun doa restu. Tanpa cinta dan doa dari kedua orang tua, skripsi ini tidak akan pernah terwujud.
2. Bapak Arif Akbarul Huda, S.Si., M.Eng selaku dosen pembimbing yang telah memberikan arahan dan bimbingan dalam pengerjaan skripsi ini.
3. Teman-teman di kelas informatika 03, yang telah saling berbagi ilmu pengetahuan dan pengalaman selama proses perkuliahan berlangsung.

Penulis juga ingin mengakui bahwa dalam skripsi ini masih terdapat kekurangan, dan penulis menghaturkan permohonan maaf yang tulus. Oleh karena itu, dengan rendah hati, penulis mengharapkan saran dan kritik dari para pembaca agar penulis dapat meningkatkan kualitas karya di masa depan. Semoga skripsi ini dapat bermanfaat bagi para pembaca dan memberikan sumbangsih positif dalam bidang ilmu pengetahuan.

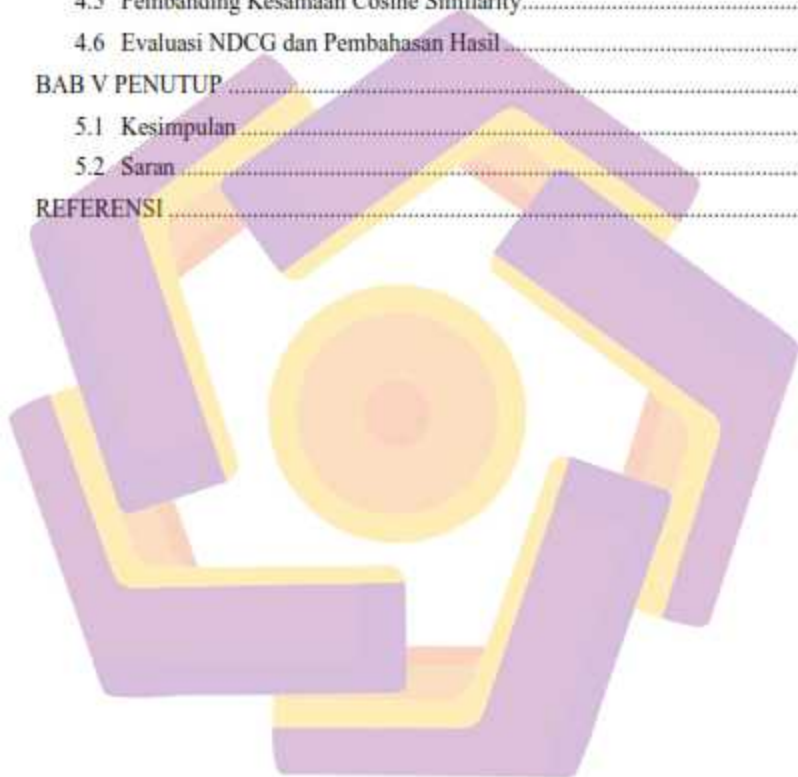
Yogyakarta, 28 Juli 2023

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vi
DAFTAR TABEL.....	viii
DAFTAR GAMBAR.....	ix
DAFTAR LAMBANG DAN SINGKATAN	x
DAFTAR ISTILAH	xi
INTISARI	xii
ABSTRACT	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	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	5
2.1 Studi Literatur.....	5
2.2 Dasar Teori	11
2.2.1 Sistem Rekomendasi	11
2.2.2 Cosine Similarity	12
2.2.3 BERT Sentence Transformers	13
2.2.4 NDCG (Normalized Discounted Cumulative Gain).....	14
BAB III METODE PENELITIAN	16
3.1 Alur Penelitian	16
3.3 Alat dan Bahan.....	20

BAB IV HASIL DAN PEMBAHASAN	21
4.1 Rancangan Arsitektur Sistem rekomendasi	21
4.2 Akuisisi Data.....	23
4.3 Data Preprocessing	25
4.4 Embedding Text dengan SBERT.....	31
4.5 Pembanding Kesamaan Cosine Similarity.....	32
4.6 Evaluasi NDCG dan Pembahasan Hasil	33
BAB V PENUTUP.....	45
5.1 Kesimpulan	45
5.2 Saran	46
REFERENSI	47



DAFTAR TABEL

Tabel 2.1. Keaslian Penelitian	7
Tabel 4.1. Sampel Data Judul Setelah Case Folding	26
Tabel 4.2. Sampel Data Content Setelah Stemming	30
Tabel 4.3. Hasil Rekomendasi indeks 0	35
Tabel 4.4. Hasil Rekomendasi indeks 1	36
Tabel 4.5. Hasil Rekomendasi Skenario 1	37
Tabel 4.6. Nilai Relevansi Skenario 1	38
Tabel 4.7. Hasil Rekomendasi Skenario 2	39
Tabel 4.8. Nilai Relevansi Skenario 2	39
Tabel 4.9. Hasil Rekomendasi Skenario 3	40
Tabel 4.10. Nilai Relevansi Skenario 3	41
Tabel 4.11. Hasil Rekomendasi Skenario 4	42
Tabel 4.12. Nilai Relevansi Skenario 4	42
Tabel 4.13. Hasil Rekomendasi Skenario 5	43
Tabel 4.14. Nilai Relevansi Skenario 5	43
Tabel 4.15. Hasil Keseluruhan Nilai NDCG	44

DAFTAR GAMBAR

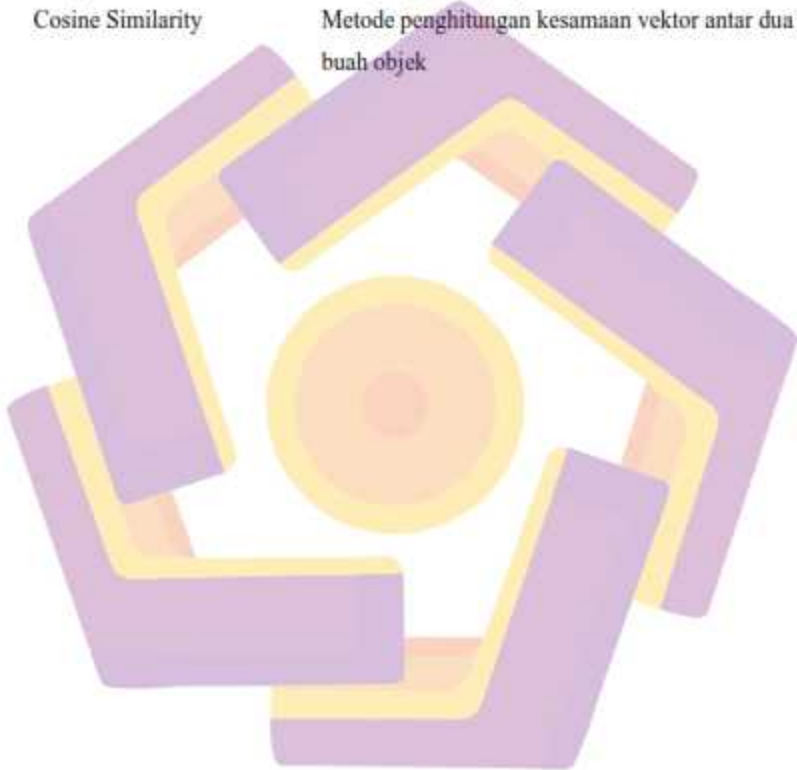
Gambar 2.1. Arsitektur Training SBERT	13
Gambar 2.2. Arsitektur Inference SBERT	14
Gambar 3.1. Alur Penelitian	16
Gambar 4.1. Arsitektur Sistem rekomendasi	22
Gambar 4.2. Potongan Kode Scrapping	23
Gambar 4.3. Dataset Artikel	24
Gambar 4.4. Kode Python Remove and Rename Columns	25
Gambar 4.5. Kode Python Remove Error Content	25
Gambar 4.6. Kode Python Case Folding	26
Gambar 4.7. Penerapan Kode Case Folding	26
Gambar 4.8. Kode Remove HTML Words	27
Gambar 4.9. Penerapan Kode Remove HTML Words	27
Gambar 4.10. Kode Remove Stop Words	28
Gambar 4.11. Sample Data Penerapan Word Tokenize	28
Gambar 4.12. Penerapan Kode Remove Stop Words	28
Gambar 4.13. Kode Stemming Words	29
Gambar 4.14. Penerapan Kode Stemming Words	29
Gambar 4.15. Arsitektur Training SBERT	31
Gambar 4.16. Kode Embedding SBERT	32
Gambar 4.17. Kode Cosine Similarity dan Sample Data	33
Gambar 4.18. Kode Pre Processing Package RSCB	34
Gambar 4.19. Kode Algorithm SBERT	34
Gambar 4.20. Kode Recommendation Articles	35

DAFTAR LAMBANG DAN SINGKATAN

$\cos \theta$	Kosinus sudut
Σ	Penjumlahan
rel_i	Nilai relevansi indeks i
NDCG	<i>Normalized Discounted Cumulative Gain</i>
DCG	<i>Discounted Cumulative Gain</i>
IDCG	<i>Ideal Discounted Cumulative Gain</i>
BERT	<i>Bidirectional Encoder Representations from Transformers</i>
SBERT	<i>Sentence Bidirectional Encoder Representations from Transformers</i>
TF-IDF	<i>Term Frequency - Inverse Document Frequency</i>
P-R	<i>Precision-Recall</i>
RSCB	<i>Recommendation System Content Based</i>

DAFTAR ISTILAH

Vektor	Besaran yang mempunyai arah
Dataset	Sekumpulan data yang terstruktur
Similarity	Nilai kemiripan antar item atau dokumen
Sampel	Sebagian jumlah populasi dari data
Cosine Similarity	Metode penghitungan kesamaan vektor antar dua buah objek



INTISARI

Sistem Rekomendasi merupakan sistem yang menganalisis item atau konten yang nantinya akan memberikan suatu rekomendasi konten atau item yang sekiranya masih dibutuhkan atau berguna bagi pengguna. Terutama dengan munculnya beragam sumber informasi online, meningkatkan kualitas rekomendasi menjadi tantangan penting untuk memberikan pengalaman pengguna yang lebih personal dan relevan. Maka dari itu pada penelitian ini penulis membuat sebuah sistem rekomendasi berbasis konten yang menggunakan dataset artikel sebagai sumber informasi untuk memberikan rekomendasi kepada pengguna.

Teknik yang dipakai pada penelitian ini mengimplementasikan pendekatan content-based dengan memanfaatkan teknik BERT sentence transformers untuk mengekstraksi fitur dan menghasilkan representasi vektor artikel. Proses selanjutnya adalah melakukan perhitungan kemiripan kosinus antara artikel menggunakan Cosine Similarity dan kemudian akan menampilkan hasil top-N rekomendasi.

Hasil dari penelitian ini menunjukkan bahwa sistem rekomendasi yang dibuat berhasil mendapatkan nilai rata-rata NDCG sebesar 0.878 pada 5 skenario yang diujikan. Nilai NDCG yang tinggi menandakan bahwa sistem rekomendasi memiliki kualitas rekomendasi yang tinggi karena mendekati angka 1.

Kata kunci: sistem rekomendasi, SBERT, content based, cosine similarity, NDCG

ABSTRACT

Recommendation System is a system that analyzes items or content to provide content or item recommendations that are deemed useful or relevant to the user. Especially with the emergence of diverse online information sources, improving the quality of recommendations becomes a crucial challenge to offer a more personalized and relevant user experience. Therefore, in this study, the author develops a content-based recommendation system that utilizes a dataset of articles as a source of information to provide recommendations to users.

The technique employed in this research implements a content-based approach by leveraging BERT sentence transformers to extract features and generate article vector representations. The subsequent process involves calculating cosine similarity between articles using Cosine Similarity, which then presents the top-N recommendation results.

The results of this study demonstrate that the developed recommendation system achieved an average NDCG score of 0.878 across the 5 tested scenarios. The high NDCG score indicates that the recommendation system has a high-quality recommendation because it approaches the value of 1.

Keyword: recommendation system, SBERT, content based, cosine similarity, NDCG