

**SISTEM REKOMENDASI CONTENT-BASED FILTERING  
MENGUNAKAN TERM FREQUENCY – INVERSE  
DOCUMENT FREQUENCY VECTOR SIMILARTY  
UNTUK REKOMENDASI ARTIKEL BERITA**

**SKRIPSI**



disusun oleh

**Rohmad Fajarudin**

**18.11.2035**

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2022**

**SISTEM REKOMENDASI CONTENT-BASED FILTERING  
MENGUNAKAN TERM FREQUENCY – INVERSE  
DOCUMENT FREQUENCY VECTOR SIMILARTY  
UNTUK REKOMENDASI ARTIKEL BERITA**

**SKRIPSI**

untuk memenuhi sebagian persyaratan  
mencapai gelar Sarjana  
pada Program Studi Informatika



disusun oleh

**Rohmad Fajarudin**

**18.11.2035**

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2022**

# PERSETUJUAN

## SKRIPSI

### **SISTEM REKOMENDASI CONTENT-BASED FILTERING MENGUNAKAN TERM FREQUENCY – INVERSE DOCUMENT FREQUENCY VECTOR SIMILARTY UNTUK REKOMENDASI ARTIKEL BERITA**

yang dipersiapkan dan disusun oleh

**Rohmad Fajarudin**

**18.11.2035**

telah disetujui oleh Dosen Pembimbing Skripsi

pada tanggal 7 April 2022

**Dosen Pembimbing,**

**Arif Akbarul Huda, S.Si., M.Eng.**

**NIK. 190302287**

# PENGESAHAN

## SKRIPSI

### SISTEM REKOMENDASI CONTENT-BASED FILTERING MENGUNAKAN TERM FREQUENCY – INVERSE DOCUMENT FREQUENCY VECTOR SIMILARTY UNTUK REKOMENDASI ARTIKEL BERITA

yang dipersiapkan dan disusun oleh

**Rohmad Fajarudin**

**18.11.2035**

telah dipertahankan di depan Dewan Penguji  
pada tanggal 20 April 2022

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Anggit Dwi Hartanto, M.Kom**  
**NIK. 190302163**

**Arifiyanto Hadinegoro, S.Kom, MT**  
**NIK. 190302289**

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

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 20 April 2022

**DEKAN FAKULTAS ILMU KOMPUTER**

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

# PERNYATAAN

## PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 3 Juni 2022

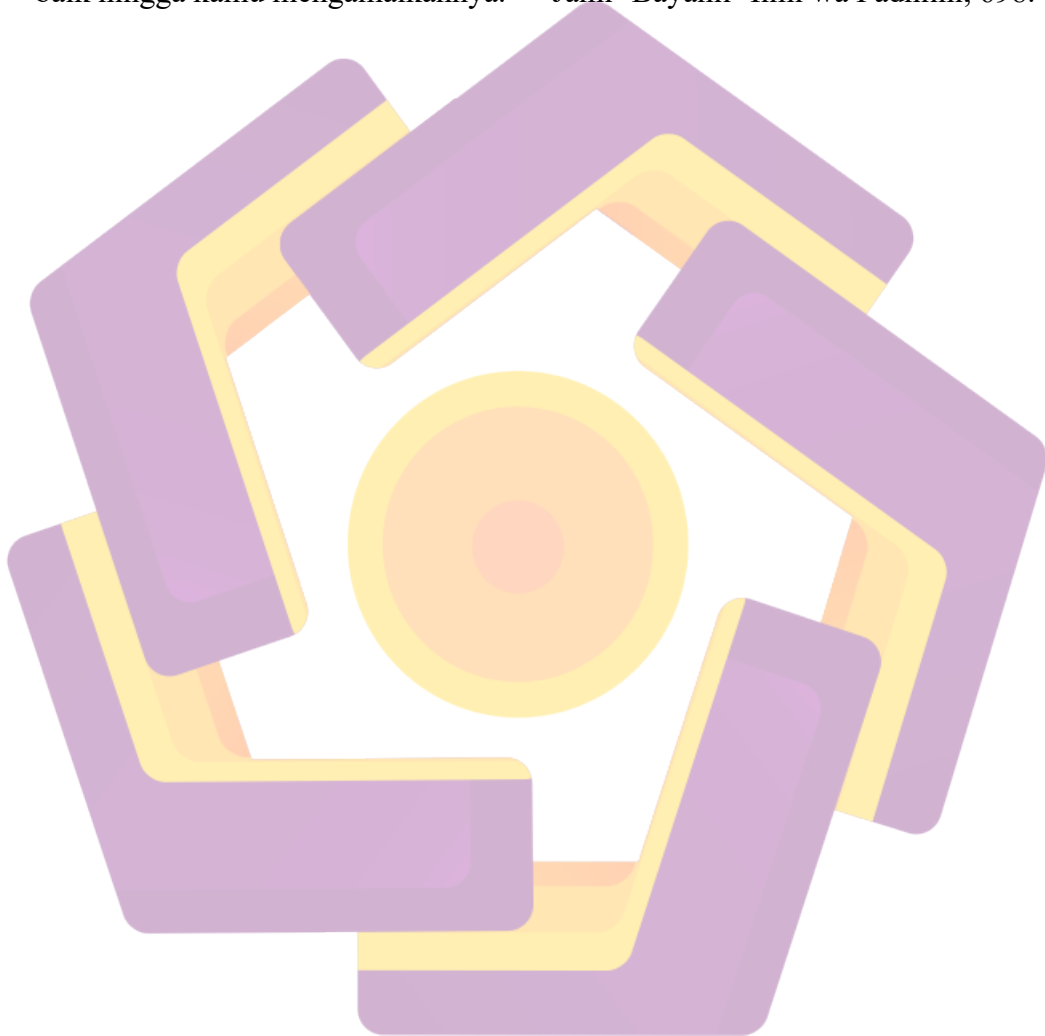


Rohmad Fajarudin  
NIM. 18.11.2035

## MOTTO

“Kesabaran tidaklah pasif, sebaliknya, kesabaran adalah kekuatan yang terkonsentrasi.” – Bruce Lee.

“Kamu tidak akan bertakwa hingga kamu berilmu; dan ilmumu tidak akan baik hingga kamu mengamalkannya.” – Jami’ Bayanil ‘Ilmi wa Fadhlihi, 698.



## PERSEMBAHAN

Laporan skripsi ini saya persembahkan kepada:

1. Kedua orang tua yang senantiasa mendukung dan mendoakan saya, Bapak Poniman Pawiro Sumarto dan (almarhumah) Ibu Sumarni.
2. Kakak-kakak dan adik-adik saya yang senantiasa memberikan banyak dukungan, dorongan dan semangat.
3. Bapak Arif Akbarul Huda, S.Si., M.Eng. yang telah memberikan kesempatan untuk bergabung kedalam Grup Riset 2021/2022 dan memberikan waktu, ilmu, dan bimbingan sehingga proses skripsi menjadi lebih lancar.
4. Semua dosen dan asisten dosen Universitas Amikom Yogyakarta yang telah mengajarkan berbagai ilmu selama kuliah.
5. Teman-teman dari 18-S1IF-04, Angkatan 2018, AMCC, dan Grup Riset 2020/2021 yang telah menemani dan menjadi teman diskusi.
6. Komunitas Kaggle, pengembang sumber terbuka (Open Source), serta para ahli lainnya yang telah membagikan ilmunya, menyediakan berbagai pustaka (*library*) dan perangkat lunak pendukung lain yang sangat bermanfaat dalam membantu jalannya riset.
7. Dan pihak-pihak lain yang tidak dapat saya tulis satu persatu yang telah berkontribusi baik secara langsung dan tidak langsung.

## KATA PENGANTAR

Assalamualaikum Wr. Wb.

Segala puja dan puji syukur kehadirat Allah SWT atas berkah dan rahmat yang telah dilimpahkan-Nya. Sholawat dan salam tetap tercurahkan kepada Rasulullah SAW yang telah mengantarkan umat ke zaman yang penuh dengan ilmu pengetahuan.

Skripsi adalah syarat bagi mahasiswa untuk menyelesaikan pendidikan Sarjana di Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Terselesaikannya skripsi ini tidak dapat terlepas dari campur tangan berbagai pihak yang telah banyak mendukung dan berkontribusi dalam proses penelitian dan penulisan baik secara langsung dan tidak langsung. Oleh sebab itu, penulis sekali lagi ingin mengucapkan rasa terima kasih kepada berbagai pihak mulai dari orang tua, kakak, adik, dosen pembimbing (Bapak Arif Akbarul Huda, S.Si., M.Eng), segenap dosen dan asisten Universitas Amikom Yogyakarta, dan pihak-pihak lain yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari masih terdapat banyak kekurangan dalam penulisan skripsi ini baik dari materi, metode, cara penyajian, ataupun aspek-aspek lain. Oleh karena itu, penulis mengharapkan banyak kritik dan saran yang membangun demi memperbaiki kekurangan yang ada.

Wassalamu'alaikum Wr. Wb.

Yogyakarta, 3 Juni 2022

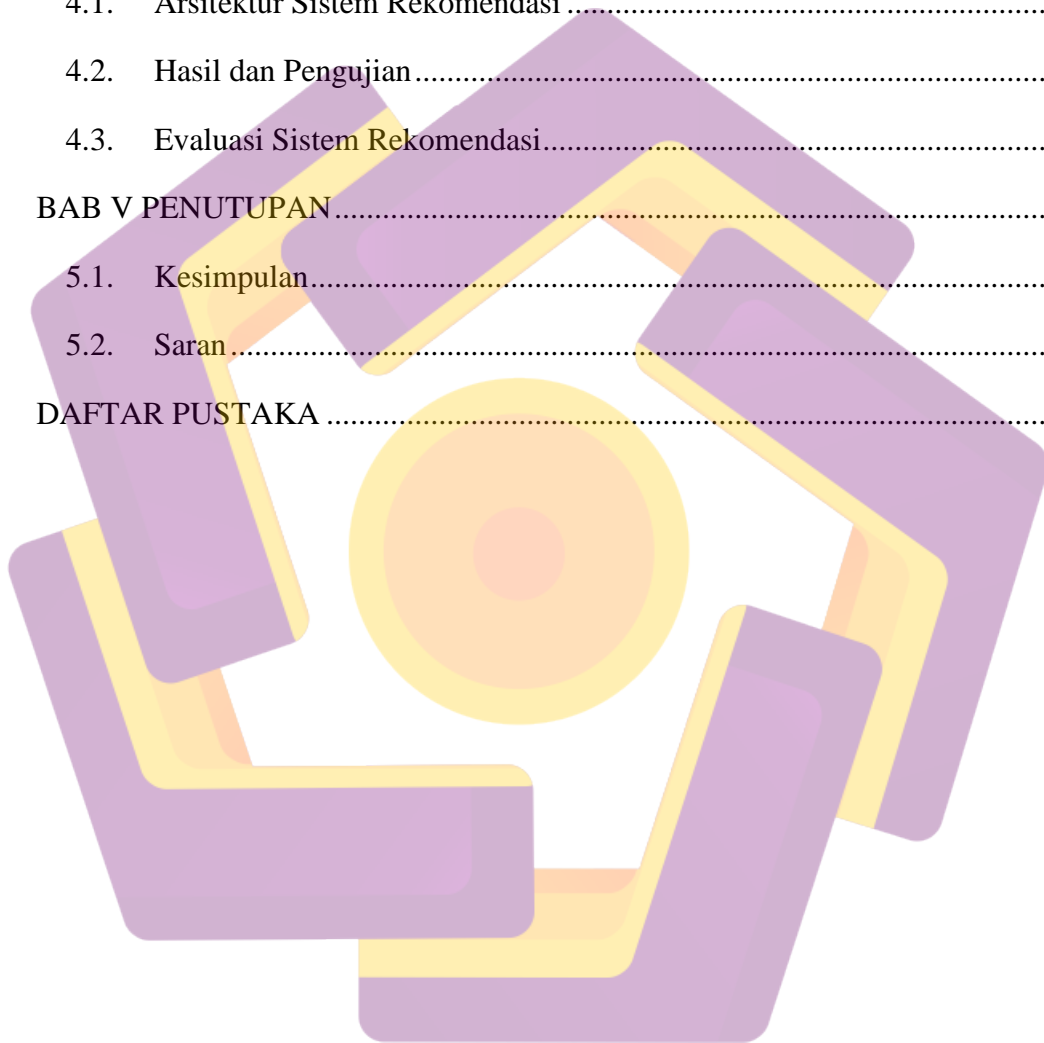
Rohmad Fajarudin



## DAFTAR ISI

COVER .....	i
PERSETUJUAN .....	ii
PENGESAHAN .....	iii
PERNYATAAN.....	iv
MOTTO .....	v
PERSEMBAHAN.....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR .....	xi
INTISARI.....	xii
ABSTRACT.....	xiii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	4
1.3. Batasan Masalah.....	4
1.4. Maksud dan Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	4
1.6. Metode Penelitian.....	5
1.7. Sistematika Penulisan.....	6
BAB II LANDASAN TEORI.....	7
2.1. Kajian Pustaka.....	7
2.2. Dasar Teori.....	9

BAB III METODE PENELITIAN.....	16
3.1. Alat dan Bahan Penelitian .....	16
3.2. Alur Penelitian.....	16
BAB IV HASIL DAN PEMBAHASAN .....	28
4.1. Arsitektur Sistem Rekomendasi .....	28
4.2. Hasil dan Pengujian.....	31
4.3. Evaluasi Sistem Rekomendasi.....	36
BAB V PENUTUPAN.....	38
5.1. Kesimpulan.....	38
5.2. Saran.....	38
DAFTAR PUSTAKA .....	39



## DAFTAR TABEL

Tabel 2.2.1. Contoh representasi data dalam <i>Vector Space Model</i> .....	13
Tabel 3.1.1. Daftar perangkat lunak dan pustaka.....	16
Tabel 3.2.1. Daftar jenis interaksi .....	20
Tabel 3.2.2. Sampel data interaksi pengguna.....	21
Tabel 3.2.3. Sampel data interaksi setelah <i>pre-processing</i> .....	24
Tabel 3.2.4. Sampel hasil kalkulasi <i>Cosine Similarity</i> pada pengguna.....	26
Tabel 4.1.1. Struktur data artikel.....	31
Tabel 4.1.2. Struktur data interaksi .....	31
Tabel 4.2.1. Hasil relevansi token untuk pengguna 1 .....	32
Tabel 4.2.2. Hasil rekomendasi artikel untuk pengguna 1 .....	33
Tabel 4.2.3. Hasil relevansi token untuk pengguna 2 .....	33
Tabel 4.2.4. Hasil rekomendasi artikel untuk pengguna 2.....	33
Tabel 4.2.5. Hasil relevansi token untuk pengguna 3 .....	34
Tabel 4.2.6. Hasil rekomendasi artikel untuk pengguna 3 .....	34
Tabel 4.2.7. Hasil relevansi token untuk pengguna 4 .....	35
Tabel 4.2.8. Hasil rekomendasi artikel untuk pengguna 4.....	35
Tabel 4.3.1. Hasil evaluasi dengan metrik <i>Recall</i> .....	36

## DAFTAR GAMBAR

Gambar 2.2.1. Contoh perhitungan <i>Cosine Similarity</i> .....	14
Gambar 3.2.1. <i>Flowchart</i> alur penelitian .....	17
Gambar 3.2.2. Data <i>RSS Feed</i> artikel dalam format XML .....	19
Gambar 3.2.3. Data <i>RSS Feed</i> artikel yang dikonveksi ke JSON .....	19
Gambar 3.2.4. Kode program untuk menghasilkan data interaksi .....	21
Gambar 3.2.5. <i>Flowchart</i> tahap perancangan .....	23
Gambar 3.2.6. <i>Pre-process</i> dan <i>Feature Engineering</i> data artikel .....	25
Gambar 4.1.1. Arsitektur sistem rekomendasi .....	28
Gambar 4.1.2. <i>Class diagram package</i> sistem rekomendasi .....	30
Gambar 4.2.1. Contoh penggunaan package untuk membuat rekomendasi .....	32
Gambar 4.3.1. Grafik evaluasi <i>Recall@5</i> .....	37
Gambar 4.3.2. Grafik evaluasi <i>Recall@10</i> .....	37

## INTISARI

Ledakan pertumbuhan dari berbagai jenis informasi yang tersedia di internet menimbulkan dampak berupa harus memilah mana yang relevan dari luasnya lautan informasi. Situs web berita adalah salah satu yang terdampak dari ledakan informasi ini.

Untuk membantu menemukan informasi yang relevan, dikembangkanlah sistem rekomendasi. Sistem rekomendasi akan membantu mengestimasi nilai prediksi atau relevansi dari sebuah artikel dan kemudian membuat *ranking* atas artikel-artikel tersebut yang sesuai dengan minat pengguna. Penelitian ini menawarkan solusi sistem rekomendasi *Content-based Filtering* dalam bentuk *package* atau pustaka pada bahasa pemrograman Python yang dapat dipergunakan untuk mempercepat pembuatan prototipe sistem.

Dengan menggunakan data dari portal berita Kabar Informatika Universitas Amikom Yogyakarta sebagai data uji, *package* sistem rekomendasi *Content-based Filtering* pada penelitian ini mampu memberikan skor *Recall@5* sekitar 69% dan *Recall@10* sekitar 82%.

**Kata Kunci:** Artikel Berita, Content-based Filtering, Cosine Similarity, Sistem Rekomendasi, TF-IDF

## **ABSTRACT**

*The explosion of many kind of information available on Internet introduced a problem where people now must filter which information is relevant to them. News website is one of many which are affected by this impact.*

*To help solve this problem, recommender system is developed. A recommender system helps to estimate the prediction value or relevancy of an article and create a ranking according to the user's taste. This research offers Content-based Filtering recommender system as a Python package library to help prototype the system quickly.*

*Using data from Kabar Informatika news portal of University of Amikom Yogyakarta as the test data, the developed Content-based Filtering recommender system is able to produce Recall@5 score at around 69% and Recall@10 at around 82%.*

**Keywords:** *Content-based Filtering, Cosine Similarity, News Article, Recommender System, TF-IDF*

