

Sistem Rekomendasi Pemilihan Program MSIB Bagi Mahasiswa Pendidikan Informatika

Dian Budi Elnursa¹, Vesy Nofriana², Agus Syamsuri³, Laili Cahyani⁴

^{1, 2, 3, 4} Pendidikan Informatika, Universitas Trunojoyo Madura, Bangkalan, Indonesia

¹200631100060@student.trunojoyo.ac.id, ²200631100045@student.trunojoyo.ac.id,

³200631100062@student.trunojoyo.ac.id, ⁴laili.cahyani@trunojoyo.ac.id

Informasi Artikel

Article history:

Diterima Juni 12, 2023

Ditinjau Juni 23, 2023

Dipublikasi Juni 30, 2023

Kata Kunci:

Recommendation System

MSIB Program

Content Based Filtering

Cosine Similarity

ABSTRACT

This study discusses a recommendation system using a content-based filtering method with cosine similarity to help informatics education students choose the right Certified Independent Study and Internship Program (MSIB). The data used is data based on student interests, program and course data available on the MSIB web portal. The content-based filtering method is used to consider the suitability between student preferences and the MSIB program curriculum, while the cosine similarity algorithm is used to calculate the similarity score between different contents. The development of this recommendation system can assist informatics education students in choosing the MSIB program that is in accordance with the preferences of the student's interest profile. The results of the system evaluation obtained an average precision level of 89.4%, indicating that the list of recommendations provided by the system is very good and very relevant according to user preferences.

*Koresponden Author:

Laili Cahyani,

Pendidikan Informatika,

Universitas Trunojoyo Madura,

Jl. Raya Telang Bangkalan Jawa Timur, Indonesia, 69162.

Email: laili.cahyani@trunojoyo.ac.id



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

1. PENDAHULUAN

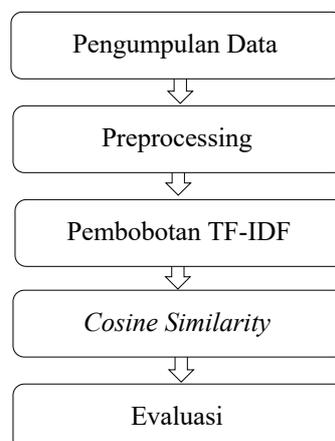
Menteri Pendidikan dan Kebudayaan (Mendikbud), Nadiem Anwar Makarim, mengubah kurikulum 2013 menjadi kurikulum MBKM (Merdeka Belajar Kampus Merdeka) pada tahun 2019. Kampus Merdeka merupakan wujud pembelajaran di perguruan tinggi yang otonom dan fleksibel sehingga tercipta kultur belajar yang inovatif, tidak mengekang, dan sesuai dengan kebutuhan mahasiswa. Transformasi pendidikan melalui kebijakan merdeka belajar merupakan salah satu langkah untuk mewujudkan SDM Unggul Indonesia yang memiliki Profil Pelajar Pancasila [3]. Salah satu bentuk kampus merdeka tersebut adalah program magang/praktik kerja dan membuat studi/proyek independent. Dengan dibukakan kesempatan tersebut, mulai banyak mahasiswa yang menginginkan pengalaman magang atau melakukan Studi Independen Bersertifikat sebagai bagian dari perkembangan kariernya. Namun, dengan banyaknya pilihan program Magang dan Studi Independen Bersertifikat yang tersedia, seringkali menjadi suatu tantangan bagi mahasiswa untuk memilih program yang sesuai dengan minat dan keahlian mereka [7].

Hal ini mendorong perlunya suatu sistem yang dapat memberikan rekomendasi program Magang dan Studi Independen Bersertifikat (MSIB) yang tepat bagi mahasiswa. Sistem rekomendasi adalah seperangkat peralatan dan teknik yang dapat memprediksi nilai preferensi atau *rating* sebuah item dengan memanfaatkan informasi personal seseorang dan karakteristik item yang ditujukan untuk memberikan rekomendasi yang efektif bagi pengguna [4]. Dalam penelitian ini, penulis membangun suatu sistem rekomendasi menggunakan *content-based filtering* dengan algoritma *cosine similarity* dan Algoritma TF-IDF untuk membantu mahasiswa dalam memilih program Magang dan Studi Independen Bersertifikat dengan tepat. *Cosine Similarity* digunakan untuk menghitung sudut yang terbentuk. Kelebihannya meskipun ada data yang mirip namun memiliki jarak yang jauh, maka sudut yang terbentuk akan kecil dan perhitungannya jauh lebih baik daripada menggunakan metode *Euclidean Distance* dan lainnya. TF-IDF digunakan karena salah satu metode *text-processing* yang simpel dan relevan [10].

Sebelumnya telah dikembangkan penelitian serupa, salah satunya pembuatan sistem rekomendasi pemilihan program studi untuk siswa SMA sederajat menggunakan *content-based filtering* dan *collaborative filtering yang memperoleh* hasil rekomendasi yang relevan dengan tingkat akurasi 85% walaupun ditemukan juga beberapa hasil yang kurang relevan akibat jumlah dataset terlalu banyak [9]. Penelitian lainnya berjudul “*Pemetaan Profil Mahasiswa Untuk Memprediksi Peminatan Mahasiswa*” juga mengembangkan sistem rekomendasi peminatan mahasiswa berdasarkan profil mahasiswa dengan hasil yang cukup baik sesuai kepribadian mahasiswa melalui metode *content-based filtering* algoritma TF-IDF [2]. Dari referensi tersebut diperoleh bahwa *content-based filtering* memiliki hasil yang cukup efektif dan relevan sebagai metode dalam sistem rekomendasi, penggantian data menjadi lebih mudah karena mesin rekomendasi tidak bergantung pada dataset yang digunakan. Selain itu, belum adanya topik pembahasan terkait sistem rekomendasi pemilihan MSIB program kampus merdeka bagi mahasiswa *urgent* saat ini. Diharapkan dengan adanya sistem rekomendasi ini, mahasiswa dapat lebih mudah menemukan program Magang dan Studi Independen Bersertifikat yang sesuai dengan minat dan keahlian mahasiswa, sehingga dapat memperoleh pengalaman yang bermanfaat dan mendukung perkembangan kariernya mahasiswa di masa depan.

2. METODE PENELITIAN/ALGORITMA

Dalam penelitian ini menggunakan metode *content-based filtering* untuk melakukan penelitian konten yang akan dijadikan parameter pembobotan pemilihan program MSIB (Magang dan Studi Independen Bersertifikat) bagi mahasiswa Pendidikan Informatika. Dalam Implementasinya *Content Based Filtering* menggunakan TF-IDF untuk melakukan pembobotan nilai atribut, sedangkan algoritma *cosine similarity* digunakan untuk menghitung skor kemiripan antar konten yang berbeda.



Gambar 2. 1 Alur Penelitian

2.1. Pengumpulan Data

Tahapan pertama yang dilakukan yaitu pengumpulan data dilakukan dengan menggunakan metode survei responden untuk mahasiswa sebagai alat pengumpulan data primer serta menggunakan metode studi pustaka pada laman program MSIB. Dataset primer adalah data kegiatan yang tersedia pada laman yakni <https://kampusmerdeka.kemdikbud.go.id/> program MSIB (Magang dan Studi Independent Bersertifikat) [3]. Sedangkan data lain yang akan digunakan untuk pembandingan konten meliputi data mahasiswa Pendidikan Informatika Universitas Trunojoyo Madura berupa preferensi profile minat bakat. Keseluruhan data akan disimpan dalam file berformat csv. Data format csv tersebut akan dikonversi menggunakan library pandas menjadi sebuah dataset sebelum dilakukan pengolahan data.

2.2. Preprocessing

Tahapan preprocessing ini data yang telah dikumpulkan akan diolah dan diproses agar dapat digunakan dalam sistem rekomendasi. Dataset akan diubah menjadi bentuk vektor yang merepresentasikan konten dari matakuliah tersebut. Proses ini dilakukan dengan menggunakan library NTLK [8]. Berikut proses preprocessing yang dilakukan:

a. Seleksi Fitur

Seleksi fitur digunakan untuk memastikan bahwa data yang digunakan memiliki informasi yang berguna dan dapat mewakili preferensi atau profil dari sebuah konten.

b. Cleaning Text

Berisi beberapa tahapan diantaranya adalah *case folding*, penghapusan nomor, simbol atau tanda baca menggunakan library NTLK. Setelah dilakukan cleaning pada data maka akan dihasilkan kumpulan kata yang merepresentasikan profil pada sebuah konten, dan kumpulan data tersebut akan digunakan untuk proses training.

c. Stopword Removal

Proses ini akan menghapus kata yang tidak memiliki makna atau kurang efektif seperti kata pada, dan, hingga, dan lain-lain.

2.3. Pembobotan TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF), pembobotan kata berfungsi untuk mengubah data berupa teks menjadi numerik matriks fitur. TF-IDF merupakan salah satu teknik yang digunakan dalam pengolahan teks untuk memberikan bobot pada kata-kata dalam sebuah dokumen. Proses ini dilakukan menggunakan library *TfidfVectorizer* dari Scikit Learn Tujuan dari TF-IDF adalah untuk mengidentifikasi kata-kata yang paling penting dalam suatu dokumen atau kumpulan dokumen. Term Frequency (TF) adalah nilai frekuensi kemunculan suatu kata dalam sebuah dokumen, sedangkan Inverse Document Frequency (IDF) yang memberikan bobot pada kata-kata yang muncul jarang di seluruh dokumen [11].

Berikut nilai TF dapat dihitung dengan menggunakan rumus:

$$TF = \frac{\text{jumlah kemunculan kata dalam dokumen}}{\text{jumlah kata dalam dokumen}} \quad (1)$$

Sedangkan, IDF dapat dihitung dengan menggunakan rumus:

$$IDF = \log \frac{N}{n} \quad (2)$$

Memperoleh alur proses transformasi dengan rumus:

$$w_{i,j} = tf_{i,j} \times \log \left(\frac{N+1}{df_i} \right) + 1 = tf \times idf \quad (3)$$

2.4. Cosine Similarity

Hasil data dari proses TF-IDF akan dihitung nilai kesamaannya dengan data konten lainnya. Perhitungan tersebut dilakukan dengan menggunakan Cosine Similarity dan bantuan library *cosine_similarity* pada scikit learn [5]. Library ini akan melakukan

perhitungan nilai kesamaan pada setiap konten berdasarkan matriks fitur yang ada dan menghasilkan output berupa nilai cosine dalam bentuk array yang nantinya akan digunakan sebagai pertimbangan untuk memberikan rekomendasi MSIB kepada mahasiswa. Berikut adalah rumus perhitungan cosine similarity [1].

$$\text{Sim}(i_1, i_2) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (4)$$

Keterangan:

- $\cos(\theta)$ = tingkat kesamaan dokumen
- A = term ke i untuk vektor dokumen ke A
- B = term ke i untuk vektor dokumen ke B
- n = jumlah data

2.5. Preprocessing

Model yang telah dibuat akan dievaluasi untuk mengukur performa kinerja oleh sistem. Evaluasi dilakukan dengan menggunakan precision. Precision untuk mengetahui tingkat ketepatan prediksi data yang benar [6]. Hasil evaluasi akan digunakan untuk menentukan apakah model sistem rekomendasi yang dibuat sudah dapat diimplementasikan secara efektif dalam program MSIB. Tingkat akurasi pada sistem diperoleh dari representasi prediksi hasil skor similarity konten pada dataset terhadap setiap konten acuan yang akan digunakan. Adapun perhitungan *precision* untuk pengevaluasian performa sistem rekomendasi:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

Keterangan:

- True Positive (TP): Jumlah prediksi relevan untuk rekomendasi yang mirip atau serupa.
- False Positive (FP): Jumlah prediksi tidak relevan tetapi hasil prediksi model relevan.

3. HASIL PENELITIAN DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data dilakukan dengan melakukan pendataan secara manual terhadap mitra MSIB yang linear dengan Program Studi Pendidikan Informatika, data mitra MSIB tersebut didapat dari website MSIB [3] dengan total sebanyak 27 data mitra yang dikumpulkan, serta 5 data mahasiswa berdasarkan preferensi profil minat. Beberapa dataset yang didapat dapat dilihat pada Tabel 1 dan Tabel 2.

Tabel 1. Dataset MSIB

partner_name	Address	description	program	course_name
PT Orbit Ventura Indonesia	Veteran RI Building 15th Floor Unit Z15-002, Plaza Semanggi Jl.Jend.Sudirman Kav. 50, Jakarta Indonesia – 12930	AI Project Cycle: AI Research Methods: Python 1 programming.AI Technology Logic and Concept: Final Project 5. Professional Ethics & Company Skills	Studi Independen	AI 4 Jobs
PT Mitra Semeru Indonesia	Jl. Mampang Prapatan Raya no 84A, Tegal	AI Technology Logic and Concept AI Project Cycle:	Studi Independen	AI for StartUps

	Parang, Mampang Prapatan, Jakarta Selatan	Python programming: AI Research Methods: Final Project: Entrepreneurship		
.....
PT Generasi Anak Muda Berkarya	Grand Slipi Tower, CHR Office Unit O Lt 9, Jl. Letjend S Parman Kav 22- 24, Jakarta Barat, 11480	Applying Design on Android Storing Data into Local Database Analyzing business processes of products Understanding Clean Code Architecture Integrating APIs	Magang	Mobile Android Development

Tabel 2. Dataset Mahasiswa Minat

name	address	minat_bidang
Mahasiswa1	Karang Anyar	AI dan Machine Learning
Mahasiswa2	Bangkalan	Design dan Game
.....
Mahasiswa15	Tangerang	Server dan Network Administrator

3.2 Preprocessing

Setelah dataset terkumpul, dilakukan tahapan *preprocessing*. Pada tahapan ini data akan diolah dan diproses agar dapat digunakan dalam sistem rekomendasi. Adapun proses preprocessing yang dilakukan, antara lain:

a. Seleksi Fitur

Data csv yg telah dikonversi menjadi dataset akan diseleksi oleh sistem. Tujuan seleksi ini adalah memastikan bahwa data yang digunakan memiliki informasi yang berguna dan dapat mewakili preferensi atau profil dari sebuah konten. Contoh data yang dapat mewakili profil konten meliputi nama mitra, program, dan kursus/posisi. Kode Sumber 1 adalah potongan kode proses seleksi fitur.

```
def print_description(index):
    data = df[df.index == index][['description', 'partner_name', 'program',
    'course_name']].values[0]
    if len(data) > 0:
        print(data[0], '\n')
        print('Nama Mitra\t\t:', data[1])
        print('Program\t\t\t:', data[2])
        print('Kursus atau Posisi\t:', data[3])
        print()
```

Kode Sumber 1. Seleksi Fitur

Hasil dari seleksi fitur ini ditampilkan seperti gambar 3.1

```
AI Technology Logic and Concept AI Project Cycle: Python programming: AI Research Methods

Nama Mitra      : PT Mitra Semeru Indonesia
Program         : Studi Independen
Kursus atau Posisi : AI for StartUps
```

Gambar 1. Hasil Proses Seleksi Fitur.

b. Cleaning Text

Setelah dilakukan seleksi fitur selanjutnya dilakukan cleaning text pada fitur *description* hingga bersih dari huruf besar, angka dan symbol atau tanda baca. Cleaning Text dilakukan dengan menggunakan kode program yang seperti Kode Sumber 3. Namun, sebelum itu diperlukan mengimport library NLTK dan juga download paket corpus stopwords dengan menggunakan kode program seperti Kode Sumber 2. Hasil Cleaning Text disimpan pada kolom *desc_clean* data seperti pada Gambar 2.

c. Stopword Removal

Pada tahapan ini, setelah fitur *description* pada dataset MSIB bersih dari huruf besar, angka dan symbol atau tanda baca selanjutnya dilakukan proses stopwords untuk menghilangkan kata hubung dengan menggunakan kode program seperti pada Tabel 2. Proses Stopword menghasilkan data seperti Gambar 3.

```
import nltk
nltk.download('stopwords')
```

Kode Sumber 2. Import NTLK dan Download Stopword

```
from nltk.corpus import stopwords
clean_spc1 = re.compile('[/(){} \[\] \@,;]')
clean_symbol = re.compile('[^0-9a-z #+ _]')
stopwords = set(stopwords.words('english'))

def clean_text(text):
    """
    text: a string
    return: modified initial string
    """
    text = text.lower() # lowercase text
    text = clean_spc1.sub('', text)
    text = clean_symbol.sub("", text)
    text = ' '.join(word for word in text.split() if word not in stopwords) # hapus stopwords
    return text

df['desc_clean'] = df['description'].apply(clean_text)
```

Kode Sumber 3. Proses Cleaning Text dan Stopword

partner_name	address	description	program	course_name	desc_clean
PT Orbit Ventura Indonesia	Veteran RI Building 15th Floor Unit Z15-002, P...	AI Project Cycle: AI Research Methods: Python ...	Studi Independen	AI 4 Jobs	ai project cycle ai research methods python 1 ...
PT Mitra Semeru Indonesia	Jl. Mampang Prapatan Raya no 84A, Tegall Parang...	AI Technology Logic and Concept AI Project Cyc...	Studi Independen	AI for StartUps	ai technology logic concept ai project cycle p...
PT Nurul Fikri Cipta Inovasi	Jl. Situ Indah No.116 RT. 006 RW. 010 Kel. Tug...	Desain Web Version Control System (VCS): Git/G...	Studi Independen	Akademi Fullstack Web Developer	desain web version control system vcs git gith...
PT Ruang Raya Indonesia	Jl. Dr. Saharjo No.161, Manggarai Selatan, Teb...	Final Project Software Engineering: Implementi...	Studi Independen	Backend Engineering	final project software engineering implementin...
PT LENTERA BANGSA BENDERANG	(Alamat Tidak Tersedia)	Know the Introduction to the World of Programm...	Studi Independen	Backend Javascript	know introduction world programming able testi...
PT DWI INTI PUTRA	(Alamat Tidak Tersedia)	Capstone Project Support Group Introduction to...	Studi Independen	Data Analyst	capstone project support group introduction py...
PT. INDOBIT DIGITAL RAYA	Educenter building unit 22218 Kav	Program Onboarding Data Collection	Studi Independen	Data Analytics for Business	program onboarding data collection

Gambar 2. Hasil Cleaning Text dan Stopword Removal

Gambar 2. menunjukkan bahwa dataset berhasil di *cleaning* yang kemudian disimpan dalam variabel *desc_clean* untuk dapat diolah pada tahapan selanjutnya TF-IDF.

3.3 Pembobotan TF-IDF

Pada tahap ini dilakukan pembobotan TF-IDF setelah dilakukan *preprocessing* data pada tahap yang telah dilakukan sebelumnya. Proses ini dilakukan menggunakan library *TfidfVectorizer* dari Scikit Learn. TF-IDF ini dilakukan untuk mengidentifikasi kata-kata yang paling penting dalam suatu dokumen atau kumpulan dokumen.

```
from sklearn.feature_extraction.text import TfidfVectorizer
df.set_index('partner_name', inplace=True)
tf = TfidfVectorizer(analyzer='word', ngram_range=(1, 3), min_df=0,
stop_words='english')
tfidf_matrix = tf.fit_transform(df['desc_clean'])
```

Kode Sumber 4. Pembobotan TF-IDF

Potongan kode program Kode Sumber 4 dapat dilihat merupakan proses yang berguna untuk melakukan transformasi data dengan menggunakan fungsi bernama *fit transform* yang ada pada *TfidfVectorizer* yang kemudian disimpan dalam variabel *vectorizer*. Hasil pembobotan TF-IDF dapat dilihat pada gambar 3.

3d modelling animation	ability	...	world programming	world programming able	write
0.00000	0.000000	...	0.000000	0.000000	0.000000
0.00000	0.000000	...	0.000000	0.000000	0.000000
0.00000	0.000000	...	0.000000	0.000000	0.000000
0.00000	0.000000	...	0.000000	0.000000	0.000000
0.00000	0.000000	...	0.000000	0.000000	0.000000
0.00000	0.000000	...	0.000000	0.000000	0.000000
0.00000	0.052968	...	0.000000	0.000000	0.052968
0.00000	0.000000	...	0.063762	0.063762	0.000000
0.00000	0.000000	...	0.000000	0.000000	0.000000
0.00000	0.000000	...	0.000000	0.000000	0.000000
0.00000	0.000000	...	0.000000	0.000000	0.000000
0.00000	0.000000	...	0.000000	0.000000	0.000000

Gambar 3. Hasil Pembobotan TF-IDF

Gambar 3. tersebut dapat dilihat merupakan representasi matriks yang dihasilkan dari proses TF-IDF yang dimana value matriks berupa bobot antara term x dan dokumen y.

3.4 Cosine Similarity

Pada tahap ini akan dilakukan perhitungan tingkat kemiripan antar konten dengan menggunakan algoritma Cosine Similarity. Perhitungan tingkat kemiripan dihitung berdasarkan nilai cosine yang didapatkan dari perbandingan perkalian antar matriks [5]. Untuk melakukan perhitungan ini, telah disediakan fungsi *cosine_similarity* dengan parameter berupa array matriks yang dapat disimpan ke dalam variabel *cos_sim* menggunakan library scikit-learn. Berikut Kode Sumber 5. adalah potongan kode untuk melakukan perhitungan cosine similarity.

```
from sklearn.metrics.pairwise import cosine_similarity
cos_sim = cosine_similarity(tfidf_matrix, tfidf_matrix)
cos_sim
```

Kode Sumber 5. Cosine Similarity

Hasil Perhitungan nilai cosine dapat dilihat pada gambar 4.

```
array([[1.          , 0.          , 0.          , 0.          , 0.          ,
        0.          , 0.01855524, 0.01428228, 0.          , 0.          ,
        0.          , 0.03285874, 0.          , 0.01574251, 0.          ,
        0.04054021, 0.05984093, 0.          , 0.          , 0.          ,
        0.          , 0.03837546, 0.          , 0.          , 0.          ,
        0.23299691, 0.          , 0.06588987],
       [0.          , 1.          , 0.          , 0.0658063 , 0.10543963,
        0.          , 0.01263689, 0.          , 0.          , 0.01713219,
        0.          , 0.          , 0.          , 0.13816715, 0.          ,
        0.          , 0.          , 0.          , 0.          , 0.          ,
        0.          , 0.          , 0.          , 0.          , 0.04688809,
        0.          , 0.17236403, 0.          ],
       [0.          , 0.          , 1.          , 0.          , 0.          ,
        0.          , 0.          , 0.          , 0.          , 0.          ,
        0.          , 0.          , 0.02152829, 0.          , 0.          ,
        0.02476628, 0.          , 0.          , 0.          , 0.02905151,
        0.          , 0.          , 0.          , 0.          , 0.07605093,
        0.          , 0.          , 0.          ],
       [0.          , 0.0658063 , 0.          , 1.          , 0.43136928,
        0.03393717, 0.02864281, 0.00630086, 0.01679918, 0.02470109,
        0.          , 0.01414322, 0.          , 0.04883729, 0.          ,
        0.          , 0.          , 0.          , 0.          , 0.          ,
        0.03609961, 0.          , 0.03440565, 0.          , 0.03013797,
        0.          , 0.02755968, 0.01698131],
       [0.          , 0.10543963, 0.          , 0.43136928, 1.          ,
        0.03625106, 0.02901145, 0.02182012, 0.01794458, 0.03289451,
        0.03963135, 0.01178764, 0.004654 , 0.06932352, 0.          ,
        0.01070798, 0.          , 0.          , 0.          , 0.          ,
        0.03856095, 0.          , 0.01721862, 0.          , 0.04828923,
        0.00945427, 0.03881995, 0.01813912],
       [0.          , 0.          , 0.          , 0.03393717, 0.03625106,
        1.          , 0.04959168, 0.05219207, 0.00408803, 0.01059836,
```

Gambar 4. Hasil Perhitungan Cosine Similarity pada TF-IDF

3.5 Evaluasi

Berdasarkan hasil perhitungan cosine similarity, diperoleh rekomendasi pengguna sesuai urutan hasil skor nilai cosine pada tahapan sebelumnya. Hasil perbandingan tingkat kemiripan tersebut diurutkan berdasarkan tingkat kemiripan yang paling tinggi akan diberikan sebagai rekomendasi kepada pengguna. Berikut adalah hasil pemberian rekomendasi:

- a. Hasil rekomendasi berdasarkan preferensi pengguna Mahasiswa2 dengan minat bidang AI dan Machine Learning. Diperoleh hasil 3 peringkat teratas yakni:
 - 1) Urutan pertama: Machine Learning Learning Path, untuk program MSIB Studi Independen pada mitra Dicoding Indonesia.
 - 2) Urutan kedua: AI Engineer, untuk program MSIB Magang pada mitra PT Nodeflux Teknologi Indonesia.
 - 3) Urutan ketiga: AI for Startup, untuk program MSIB Studi Independen pada mitra PT Mitra Semeru Indonesia.

```
1 Mahasiswa2 Karang Anyar AI, Machine Learning
recommendations('Mahasiswa2')
Daftar Rekomendasi
-----
Mitra : Dicoding Indonesia
Program : Studi Independen
Kursus : Bangkit Academy Google, GoTo, Traveloka - Machine Learning Learning Path
-----
Mitra : PT Nodeflux Teknologi Indonesia
Program : Magang
Posisi : AI Engineer
-----
Mitra : PT Mitra Semeru Indonesia
Program : Studi Independen
Kursus : AI for Startups
-----
Mitra : PT Orbit Ventura Indonesia
Program : Studi Independen
Kursus : AI 4 Jobs
-----
```

Gambar 1. Hasil Rekomendasi Mahasiswa2

- b. Hasil rekomendasi berdasarkan preferensi pengguna Mahasiswa1 dengan minat bidang Design dan Game. Diperoleh hasil 3 peringkat teratas yakni:
 - 1) Urutan pertama: Game Development, untuk program MSIB Studi Independen pada mitra PT Kinema Systrans Multimedia.
 - 2) Urutan kedua: UI/UX Design, untuk program MSIB Studi Independen pada mitra PT Impactbyte Teknologi Edukasi.
 - 3) Urutan ketiga: Back-end Developer, untuk program MSIB Magang pada mitra PT Telkom Indonesia (Persero) Tbk.

```
0 Mahasiswa1 Bangkalan Design, Game
```

```
recommendations('Mahasiswa1')
Daftar Rekomendasi
-----
Mitra : PT Kinema Systrans Multimedia
Program : Studi Independen
Kursus : Game Development
=====
Mitra : PT Impactbyte Teknologi Edukasi
Program : Studi Independen
Kursus : UI/UX Design
=====
Mitra : PT Telkom Indonesia (Persero) Tbk
Program : Magang
Posisi : Back-end Developer
=====
Mitra : PT GIT SOLUTION
Program : Magang
Posisi : Android Developer
=====
```

Gambar 2. Hasil Rekomendasi Mahasiswa 1

- c. Hasil rekomendasi berdasarkan preferensi pengguna Mahasiswa3 dengan minat bidang Server dan Network Administrator. Diperoleh hasil 3 peringkat teratas yakni:
- 1) Urutan pertama: System Administrator - IBM AI & Cybersecurity, untuk program MSIB Studi Independen pada mitra PT Kinema Systrans Multimedia.
 - 2) Urutan kedua: Full-Stack Developer, untuk program MSIB Magang pada mitra PT Bejana Investidata Globalindo.
 - 3) Urutan ketiga: Android Developer, untuk program MSIB Magang pada mitra PT GIT SOLUTION

```
2 Mahasiswa3 Tangerang Server, Network Administrator
recommendations('Mahasiswa3')
Daftar Rekomendasi
-----
Mitra : PT Kinema Systrans Multimedia
Program : Studi Independen
Kursus : System Administrator - IBM AI & Cybersecurity
=====
Mitra : PT Bejana Investidata Globalindo
Program : Magang
Posisi : Full-Stack Developer
=====
Mitra : PT GIT SOLUTION
Program : Magang
Posisi : Android Developer
=====
Mitra : PT Hacktivate Teknologi Indonesia
Program : Studi Independen
Kursus : Golang for Back End Programmer
=====
```

Gambar 3. Hasil Rekomendasi Mahasiswa 3

Dari hasil perekomendasi tersebut, dilakukan uji coba evaluasi sistem untuk membandingkan tingkat kemiripan antar konten dataset dengan preferensi minat pengguna. Dihitung secara manual dengan melihat jumlah konten rekomendasi yang relevan dan tidak relevan. Evaluasi dilakukan dengan menggunakan precision. Untuk menghitung precision pada hasil rekomendasi dengan dilakukan pengambilan 10 sampel yang nilai precisionnya akan dijumlahkan kemudian dihitung rata-ratanya. Sistem berhasil merekomendasikan konten program MSIB relevan sesuai preferensi profil minat pengguna maka nilai precision semakin tinggi. Hasil pengujian sistem dapat dilihat pada tabel 3.8.

Tabel 3. Hasil Pengujian Sistem Menggunakan Precision

No.	Pengguna	Preferensi profil pengguna (minat)	Precision
1.	Mahasiswa1	Design, Game	100%
2.	Mahasiswa2	Ai, Machine Learning	100%
3.	Mahasiswa3	Server, Network Administrator	80%
4.	Mahasiswa4	Game, Network Administrator	77%
5.	Mahasiswa5	WEB, AI, Data Analyst	83%
6.	Mahasiswa6	Cloud Computing, Web	83%
7.	Mahasiswa7	Database System, Web, Design	80%
8.	Mahasiswa8	Design, Animation	91%
9.	Mahasiswa9	Data Analyst, Tensorflow	100%
10	Mahasiswa10	AI, Rest Api	100%
SKOR RATA-RATA			89,4%

Pengujian dari sistem rekomendasi MSIB menggunakan metode content-based filtering menghasilkan rata-rata precision sebesar 89,4%. Nilai rata-rata tersebut menunjukkan bahwa sistem rekomendasi MSIB menggunakan metode content-based filtering dapat merekomendasikan item sesuai dengan preferensi pengguna.

4. KESIMPULAN

Dari jumlah dataset yang terkumpul tidak terlalu banyak, tetapi dataset memiliki fitur yang lengkap, sistem dapat secara efektif memberikan rekomendasi program MSIB yang sesuai dengan minat mahasiswa pengguna. Hasil evaluasi sistem menunjukkan tingkat *precision* rata-rata 89,4% untuk 10 sampel uji coba pengguna, menunjukkan bahwa daftar rekomendasi yang diberikan oleh sistem sangat layak dan sangat relevan sesuai preferensi pengguna. Meski begitu, keterbatasan data dapat diatasi dengan menambah, menggabungkan, atau menggantinya. Keuntungan tersebut menjadikan penggantian data lebih mudah karena mesin rekomendasi tidak bergantung pada dataset yang digunakan. Tentunya masih banyak aspek penelitian ini yang dapat diperbaiki untuk meningkatkan kinerja sistem rekomendasi.

UCAPAN TERIMA KASIH

Peneliti mengucapkan terimakasih kepada pihak Program Studi Pendidikan Informatika Universitas Trunojoyo Madura khususnya dosen matakuliah Text Mining yang telah memberikan dukungan dan fasilitas kepada tim peneliti untuk dapat melaksanakan penelitian ini.

DAFTAR PUSTAKA

- [1] Anugrah, I. G. Penerapan Metode N-Gram dan Cosine Similarity Dalam Pencarian Pada Repositori Artikel Jurnal Publikasi. *Building of Informatics Technology and Science (BITS)*. 2021; 3(3): 275–284. <https://doi.org/10.47065/bits.v3i3.1058>
- [2] Ar-Rasyid, H., Pane, S. F., & Setyawan, M. Y. H. Pemetaan Profil Mahasiswa Untuk Memprediksi Peminatan Mahasiswa. *PETIR: Jurnal Pengkajian dan Penerapan Teknik Informatika*. 2023; 16(1). <https://doi.org/10.33322/petir.v16i1.1337>
- [3] Kemdikbud.go.id. Program Kampus Merdeka. 2022. Diakses pada 31 Mei 2023, dari <https://kampusmerdeka.kemdikbud.go.id/program>
- [4] Pavan Kumar, P., Vairachilai, S., Potluri, S., & Nandan Mohanty, S. (Eds.). *Recommender Systems: Algorithms and Applications* (1st ed.). *CRC Press*. 2021. <https://doi.org/10.1201/9780367631888>
- [5] Pavitha, N., Pungliya, V., Raut, A., Bhonsle, R., Purohit, A., Patel, A., & Shashidhar, R. Movie Recommendation and Sentiment Analysis Using Machine Learning. *Global Transitions Proceedings*. 2022: 279-284.

- [6] Putra, A. I., & Santika, R. R. Implementasi Machine Learning dalam Penentuan Rekomendasi Musik dengan Metode Content-Based Filtering. *Edumatic: Jurnal Pendidikan Informatika*. 2020; 4(1): 121-130.
- [7] Pipin, S. J., & Kurniawan, H. Analisis Sentimen Kebijakan MBKM Berdasarkan Opini Masyarakat di Twitter Menggunakan LSTM. 2022; 23(2): 197-208.
- [8] Ray, B., Garain, A., & Sarkar, R. An Ensemble-Based Hotel Recommender System Using Sentiment Analysis and Aspect Categorization of Hotel Reviews. *Applied Soft Computing*. 2021; 98: 106-935.
- [9] Rizky, M. I., Asror, I., & Murti, Y. R. Sistem Rekomendasi Program Studi untuk Siswa SMA Sederajat Menggunakan Metode Hybrid Recommendation dengan Content Based Filtering dan Collaborative Filtering. *eProceedings of Engineering*. 2020: 2776.
- [10] Raharjo, P. N., Handojo, A., & Juwiantho, H. Sistem Rekomendasi Content Based Filtering Pekerjaan dan Tenaga Kerja Potensial menggunakan Cosine Similarity. *Jurnal Invra*. 2022; 10(2): 47-56. <https://doi.org/10.48550/arXiv.1906.00041>
- [11] Wati, R., Ernawati, S., & Rachmi, H. Pembobotan TF-IDF Menggunakan Naïve Bayes Pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH TF-IDF Weighting Using Naïve Bayes on Public Sentiment on The Issue of Rising BIPIH. *Jurnal Manajemen Informatika (JAMIKA)*. 2023; 13(1): 84-93. <https://doi.org/10.34010/jamika.v13i1.9424>