

INTEGRASI *NAIVE BAYES* DAN *ITEM-BASED* COLLABORATIVE FILTERING DALAM SISTEM PEMETAAN KOMPETENSI MAHASISWA

Dini Nurmalasari^{1*}, Mardhiah Fadhli¹, Yuli Fitriasia¹, Heri R Yuliantoro²

Teknologi Rekayasa Komputer, Politeknik Caltex Riau, Indonesia, Pekanbaru 28265¹

Akuntansi Perpajakan, Politeknik Caltex Riau, Indonesia, Pekanbaru 28265, Indonesia²

dini@pcr.ac.id*, mardhiah@pcr.ac.id, uli@pcr.ac.id, heriry@pcr.ac.id

**Penulis Koresponden*

ABSTRAK

Persiapan portofolio yang unggul menjadi aspek krusial bagi mahasiswa dalam menghadapi dunia kerja, salah satunya melalui partisipasi dalam kompetisi. Namun, proses pemilihan lomba yang sesuai dengan kompetensi mahasiswa masih menjadi tantangan, mengingat banyaknya informasi lomba, keragaman minat dan kemampuan mahasiswa, serta keterbatasan dalam alokasi biaya, waktu, dan sumber daya. Penelitian ini membangun sistem rekomendasi berbasis *Hybrid Recommendation System* yang dirancang untuk memetakan kompetensi mahasiswa terhadap jenis lomba yang relevan. Sistem ini menggabungkan metode *Naive Bayes* untuk mengklasifikasikan kompetensi mahasiswa dan *Item-Based Collaborative Filtering* untuk menghitung kemiripan antar jenis lomba berdasarkan preferensi pengguna lain. Pengembangan sistem dilakukan secara bertahap menggunakan pendekatan *waterfall*, mulai dari perencanaan, analisis, desain, implementasi, hingga pengujian. Model dibangun mengikuti alur standar machine learning, meliputi pengumpulan data, eksplorasi dan pra-pemrosesan data, pembangunan model, evaluasi performa, serta integrasi metode. Data penelitian diperoleh dari profil mahasiswa, kompetensi, dan preferensi lomba yang dikumpulkan melalui survei dan database internal. Hasil evaluasi menunjukkan bahwa sistem mampu memberikan rekomendasi lomba yang relevan dengan tingkat akurasi mencapai 70%. Hasil pengujian ini menunjukkan adanya kontribusi sistem dalam membantu mahasiswa memilih lomba yang sesuai dengan kompetensinya, dengan penyajian yang mudah digunakan dalam bentuk sistem berbasis web. Selain itu hasil penelitian ini memberikan kontribusi berupa kemampuan sistem dalam menyederhanakan proses pencocokan kompetensi dan preferensi terhadap peluang lomba secara otomatis, sehingga dapat meningkatkan partisipasi mahasiswa dalam kegiatan yang mendukung pengembangan portofolio unggul secara lebih terarah dan efisien.

Kata kunci: *Item-Based Collaborative Filtering, Naive Bayes, Pemetaan Kompetensi Mahasiswa*

ABSTRACT

Preparing a strong portfolio is a crucial aspect for students in entering the workforce, one of which can be achieved through participation in various competitions. However, selecting competitions that align with student competencies remains a challenge due to the abundance of competition information, diversity in student interests and abilities, and limitations in budget, time, and resources. This study develops a recommendation system based on a Hybrid Recommendation System designed to map student competencies to relevant competition types. The system integrates the Naive Bayes method to classify student competencies and Item-Based Collaborative Filtering to calculate similarities between competition types based on other users' preferences. The system is developed incrementally using the waterfall approach, including the stages of planning, analysis, design, implementation, and testing. The model follows standard machine learning workflows, comprising data collection, exploration and preprocessing, model building, performance evaluation, and method integration. The research data includes student profiles, competencies, and competition preferences collected through surveys and internal databases. Evaluation results indicate that the system successfully provides relevant competition recommendations with an accuracy rate of 70%. These results demonstrate the system's contribution in assisting students to select competitions that match their competencies, presented in a user-friendly web-based application.

Keywords: *Item-Based Collaborative Filtering, Naive Bayes, Student Competency Mapping*

Histori Artikel

Diserahkan: 30 Apr 2025

Diterima setelah Revisi: 27 Mei 2025

Diterbitkan: 12 Juni 2025

1. PENDAHULUAN

Menurut Paulsonm 1991 portofolio adalah sekumpulan kegiatan siswa yang menunjukkan kecakapan, usaha dan perkembangan dalam suatu bidang atau kegiatan [1]. Salah satu fungsi dari portofolio adalah sebagai salah satu sarana penilaian yang dapat membuktikan proses dan kompetensi seorang siswa, sehingga sebuah portofolio dapat dijadikan sebagai dasar pengujian, prestasi dan karya dari seseorang. Memiliki portofolio yang baik akan sangat membantu dalam pencapaian prestasi berikutnya seperti dalam mencari pekerjaan, mendapatkan beasiswa atau melanjutkan studi bagi mahasiswa. Salah satu isi dari portofolio yang sebaiknya dipersiapkan adalah keaktifan dalam mengikuti lomba dan menghasilkan karya yang bermanfaat [2].

Saat ini informasi mengenai lomba sangat mudah didapatkan baik dari instansi pemerintah maupun organisasi dan lainnya, baik nasional maupun internasional. Namun proses pemetaan atau pemilihan keikutsertaan seorang mahasiswa dalam lomba tertentu akan menjadi hal yang kompleks dikarenakan banyaknya faktor yang mempengaruhi seperti : pemerataan keikutsertaan mahasiswa, alokasi dana, waktu pelaksanaan, optimasi sumber daya serta analisis kesiapan mahasiswa agar mendapatkan hasil yang memuaskan. Seiring dengan semakin banyaknya metode machine learning yang dapat digunakan untuk menghasilkan system rekomendasi agar memudahkan proses optimasi, maka pada penelitian ini akan dibuat sebuah system rekomendasi pemetaan kompetensi mahasiswa dalam mengikuti lomba, agar dapat mengurangi kompleksitas pemilihan kandidat peserta lomba serta mengoptimalkan potensi mahasiswa. Metode yang akan digunakan untuk melakukan proses rekomendasi adalah hybrid recommendation yaitu metode yang menggabungkan beberapa metode rekomendasi yang ada menjadi satu. Pada penelitian ini metode yang akan digunakan adalah algoritma Naïve Bayes untuk mengelompokkan jenis lomba berdasarkan kompetensi siswa dan *Item-Based* collaborative filtering untuk menghitung kemiripan jenis lomba yang dipilih mahasiswa [3]. Tujuan dari hybrid recommendation ini adalah untuk saling melengkapi agar hasil rekomendasi lebih tepat dan akurat.

Beberapa kriteria yang akan digunakan dalam pembuatan model rekomendasi dalam system ini adalah nilai akademik, minat, profil siswa dan kriteria peserta lomba untuk pengelompokan lomba. Sedangkan data yang akan digunakan untuk *Item-Based* collaborative filtering adalah kriteria lomba dengan cara memberikan rating dan preferensi dari data sebelumnya untuk mengetahui kemiripan antar jenis lomba. Setelah itu akan dilakukan perhitungan secara hybrid yang hasilnya berupa rekomendasi lomba untuk setiap mahasiswa. Teknik hybrid yang akan digunakan adalah Teknik weighted yaitu dengan cara melakukan pemberian bobot pada masing-masing metode yang digunakan. Bobot yang akan diterapkan mengikuti penelitian sebelumnya yang telah dilakukan oleh [4][5], dimana bobot untuk naïve bayes sebesar 30% dan bobot untuk *Item-Based* sebesar 70%.

2. TINJAUAN PUSTAKA

Penelitian sebelumnya mengenai pemanfaatan metode hybrid recommendation telah dilakukan untuk beberapa objek dan permasalahan yang berbeda. Diantaranya penelitian yang dilakukan oleh [6] mengenai hybrid recommendation menunjukkan peningkatan akurasi mencapai 85.5%. Penelitian serupa dilakukan dengan objek pada pemesanan hidangan oleh [7] [8] menghasilkan Tingkat presisi sebesar 80.73% dan recall 76.52%. Pada tahun 2024 [9] juga melakukan penelitian dengan memanfaatkan hybrid recommendation untuk peminjaman buku perpustakaan dan menghasilkan nilai MRR dan MAP yang lebih baik dibandingkan metode tanpa hybrid dengan nilai sebesar 0.2113 dan 0.0988. Penelitian lainnya juga dilakukan oleh [5] dalam membuat sistem rekomendasi kegiatan ekstrakurikuler siswa dengan akurasi sebesar 74%. Sementara penggunaan *Naive Bayes* dalam memberikan rekomendasi telah dilakukan pada penelitian [10][11].

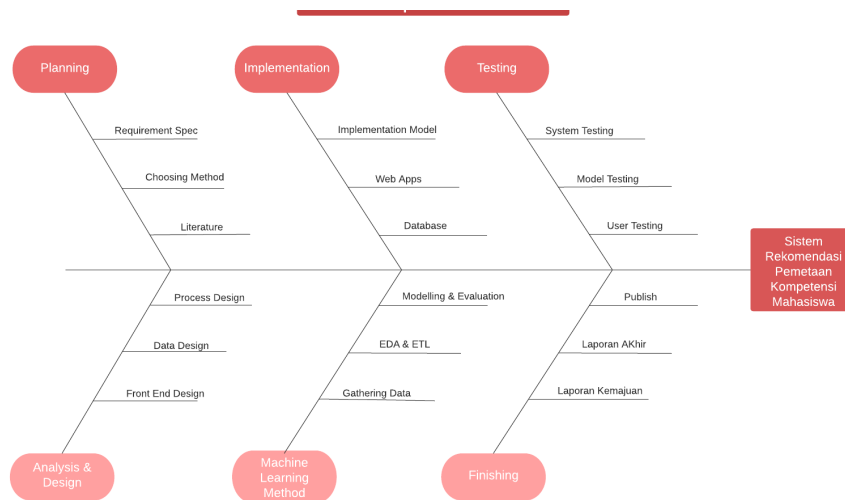
Meskipun telah banyak penelitian yang mengkaji sistem rekomendasi dalam berbagai bidang, penerapannya secara khusus untuk membantu mahasiswa dalam memilih kompetisi yang sesuai dengan kompetensi individu masih sangat terbatas. Sebagian besar sistem yang ada hanya berfokus pada rekomendasi berbasis minat tanpa mempertimbangkan dimensi kompetensi mahasiswa secara menyeluruh. Selain itu, integrasi antara pendekatan klasifikasi kompetensi dan kemiripan jenis lomba dalam satu sistem terpadu jarang ditemukan dalam literatur. Kesenjangan inilah yang mendorong

perlunya pengembangan sistem rekomendasi yang tidak hanya mempertimbangkan preferensi pengguna, tetapi juga mampu memetakan kompetensi secara sistematis agar dapat memberikan rekomendasi lomba yang lebih relevan dan terarah.

Berdasarkan paparan permasalahan yang telah disampaikan serta didukung penelitian sebelumnya mengenai pemafaatan hybrid recommendation, maka pada penelitian ini akan dikembangkan system rekomendasi pemetaan kompetensi mahasiswa dalam mengikuti lomba dengan menggunakan hybrid recommendation. Tujuan dari penelitian ini tidak hanya membantu mahasiswa dalam menentukan kegiatan atau kompetensi yang sesuai dengan kemampuan dan minat mereka, namun juga dapat mendukung portofolio yang baik bagi mahasiswa. Serta membantu management institusi Pendidikan dalam mengoptimalkan potensi mahasiswa secara lebih efektif dan efisien.

3. METODE

Metode yang digunakan dalam penelitian ini mengacu pada metode SDLC (Software Development Life Cycle) [4] serta tahapan Machine Learning untuk melengkapi proses rekomendasi system. Tahapan yang akan dilakukan adalah sebagai berikut (gambar 1) [12]:



Gambar 1. Metode Penelitian

i) Planning

Pada tahapan ini akan dilakukan perencanaan yang terkait dengan kegiatan requirement specification yaitu melakukan identifikasi kebutuhan pengguna[13]. Dalam hal ini pengguna yang akan dilibatkan adalah pengelola program studi dan lomba, serta mahasiswa sebagai pelaksana lomba. Proses pengumpulan kebutuhan pengguna akan dilakukan dengan cara wawancara, survey dan pengamatan secara langsung. Selain itu pada tahap ini juga akan dilakukan pengumpulan data, studi literatur serta persiapan perkakas yang akan digunakan. Jumlah data training yang akan dikumpulkan minimal 200 data dengan sebaran data mahasiswa dari JTI.

ii) Analysis & Design

Dari hasil identifikasi data dan identifikasi kebutuhan pengguna, selanjutnya akan dilakukan analisis dan perancangan secara teknis. Beberapa data yang akan digunakan diantaranya data profil mahasiswa, data nilai akademik, data dan informasi lomba, serta data lainnya. Selanjutnya akan dilakukan EDA dan ETL untuk melihat karakteristik dari data [14]. Proses pelabelan data akan dilakukan dengan menggunakan metode preferensi pengguna melalui survei dan wawancara dari data kompetensi mahasiswa dan keikutsertaan lomba. Selain itu pada tahapan ini juga akan dilakukan perancangan proses dengan menggunakan usecase diagram dan class diagram, serta perancangan wireframe untuk system yang akan dibuat.

iii) Implementation dan Machine Learning Method

Tahapan selanjutnya adalah melakukan implementasi terhadap perancangan yang telah dibuat. Beberapa tahap yang akan dilakukan adalah membuat pemodelan data dengan naïve bayes dan item-base collaborative filtering, kemudian memberikan bobot dan melakukan perhitungan hybrid

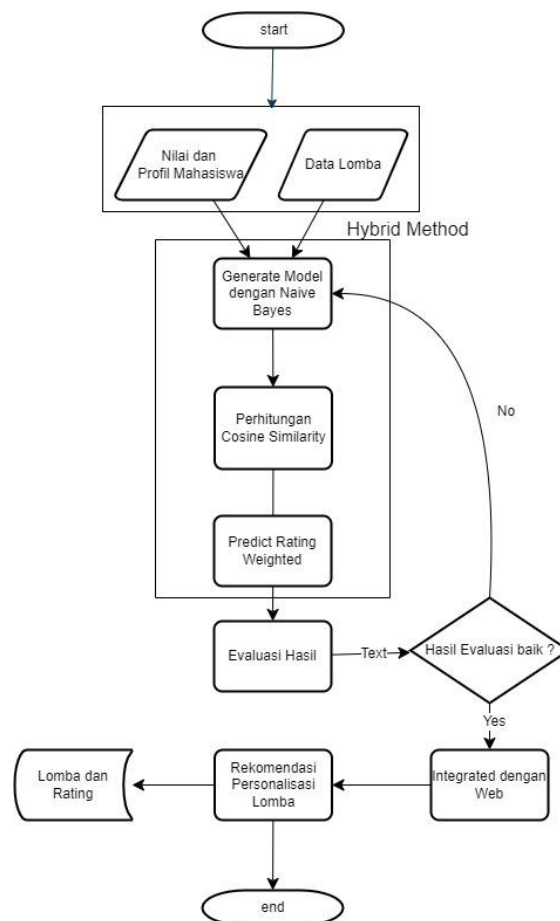
[15][16]. Setelah itu model yang didapatkan akan diintegrasikan kedalam sebuah system berbasis web yang akan dibangun dengan menggunakan Bahasa pemrograman HTML, JavaScript dan PHP, serta dibantu dengan API Flask untuk mengintegrasikan model data rekomendasi yang sudah dibuat . Label data yang akan digunakan pada proses klasifikasi data adalah jenis lomba.

iv) Testing

Pengujian merupakan tahapan system yang telah selesai dibangun. Uji coba dilakukan dengan menggunakan evaluasi tingkat Akurasi pada naïve bayes yang dilihat dari banyaknya hasil data naïve bayes yang relevan dengan dataset. Perhitungan ini bertujuan untuk melihat seberapa sesuai data rating yang diberikan *Item-Based* collaborative filtering dengan rating yang diberikan oleh mahasiswa [17] [18].

4. PERANCANGAN

Secara umum system rekomendasi yang dibangun menghasilkan keluaran berupa rekomendasi lomba yang tepat untuk setiap mahasiswa. Proses yang dilakukan pada penelitian ini terdiri dari 4 bagian besar yaitu pengolahan data, melakukan generate model dengan hybrid recommendation, mengintegrasikan dengan system berbasis web serta melakukan evaluasi dan pengujian system. Berikut (gambar 2) adalah gambaran umum dari system rekomendasi yang dibuat:



Gambar 2. Flowchart Sistem Rekomendasi

4.1 PENGOLAHAN DATA

Pada tahapan ini telah dilakukan perencanaan yang terkait dengan kegiatan requirement specification yaitu melakukan identifikasi kebutuhan pengguna. Dalam hal ini pengguna yang akan dilibatkan adalah pengelola program studi dan lomba, serta mahasiswa sebagai pelaksana lomba. Proses pengumpulan kebutuhan pengguna akan dilakukan dengan cara wawancara, survey dan pengamatan secara langsung. Selain itu pada tahap ini juga akan dilakukan pengumpulan data, studi literatur serta persiapan perkakas yang akan digunakan. Jumlah data training yang akan dikumpulkan minimal 200 data dengan sebaran data mahasiswa dari JTI. Pada gambar 3 berikut merupakan sebagian dataset yang digunakan pada penelitian ini.

mahasiswa_id	prodi	semester	ipk	kompetensi_1	kompetensi_2	minat	lomba	bidang_lomba	tingkat_lomba	jenis_lomba	rating_partisipasi
MHS001	Teknik Komputer	8	3.33	Desain UI/UX	Data Science	Web Development	Hackathon	Bisnis	Nasional	Individu	2
MHS002	Teknik Komputer	7	3.14	Jaringan	Desain UI/UX	IoT	UI/UX Competition	Teknologi	Nasional	Tim	2
MHS003	Teknik Komputer	5	3.49	Pemrograman	Basis Data	IoT	Data Challenge	Teknologi	Lokal	Tim	2
MHS004	Teknik Informatika	3	3.08	Desain UI/UX	Data Science	Web Development	Data Challenge	Keamanan	Nasional	Individu	1
MHS005	Teknik Informatika	4	3.55	Jaringan	Basis Data	IoT	Data Challenge	Keamanan	Lokal	Individu	2
MHS006	Teknik Informatika	1	3.82	Basis Data	Desain UI/UX	Desain	UI/UX Competition	Desain	Internasional	Individu	4
MHS007	Sistem Informasi	6	3.27	Desain UI/UX	Data Science	Cybersecurity	CTF	Desain	Lokal	Tim	5
MHS008	Teknik Informatika	4	3.7	Basis Data	Desain UI/UX	Web Development	CTF	Bisnis	Nasional	Tim	3
MHS009	Teknik Informatika	1	2.53	Basis Data	Desain UI/UX	Cybersecurity	Startup Pitch	Keamanan	Lokal	Tim	1
MHS010	Sistem Informasi	2	3.94	Data Science	Pemrograman	Cybersecurity	Data Challenge	Bisnis	Nasional	Tim	5
MHS011	Sistem Informasi	6	2.76	Basis Data	Desain UI/UX	IoT	CTF	Data	Lokal	Individu	4
MHS012	Teknik Komputer	7	3.1	Keamanan Siber	Basis Data	Cybersecurity	UI/UX Competition	Keamanan	Lokal	Tim	2
MHS013	Teknik Informatika	3	3.26	Jaringan	Basis Data	Web Development	UI/UX Competition	Keamanan	Internasional	Tim	4
MHS014	Teknik Informatika	7	3.96	Keamanan Siber	Data Science	Desain	CTF	Desain	Nasional	Individu	2
MHS015	Teknik Informatika	6	3.6	Basis Data	Pemrograman	AI	UI/UX Competition	Data	Lokal	Individu	1
MHS016	Teknik Komputer	4	3.5	Data Science	Pemrograman	Cybersecurity	Startup Pitch	Keamanan	Nasional	Tim	5
MHS017	Sistem Informasi	2	3.97	Desain UI/UX	Data Science	AI	CTF	Bisnis	Lokal	Tim	2
MHS018	Teknik Informatika	4	3.32	Keamanan Siber	Desain UI/UX	Web Development	CTF	Teknologi	Internasional	Individu	2
MHS019	Teknik Komputer	7	3.77	Pemrograman	Basis Data	IoT	CTF	Data	Internasional	Tim	4
MHS020	Teknik Informatika	1	3.19	Jaringan	Desain UI/UX	Desain	Data Challenge	Bisnis	Internasional	Individu	2
MHS021	Teknik Informatika	5	3.04	Basis Data	Data Science	Web Development	Data Challenge	Teknologi	Lokal	Individu	5
MHS022	Teknik Informatika	8	3.55	Keamanan Siber	Jaringan	AI	Hackathon	Data	Nasional	Individu	2
MHS023	Sistem Informasi	4	3.56	Basis Data	Data Science	AI	Startup Pitch	Keamanan	Lokal	Individu	5
MHS024	Sistem Informasi	6	3.62	Data Science	Pemrograman	IoT	UI/UX Competition	Bisnis	Nasional	Individu	2
MHS025	Sistem Informasi	8	3.0	Desain UI/UX	Data Science	Desain	Hackathon	Desain	Internasional	Tim	5
MHS026	Teknik Informatika	2	2.94	Pemrograman	Basis Data	IoT	UI/UX Competition	Data	Internasional	Individu	5
MHS027	Teknik Informatika	6	2.99	Pemrograman	Basis Data	AI	UI/UX Competition	Bisnis	Nasional	Tim	3
MHS028	Sistem Informasi	6	3.97	Keamanan Siber	Pemrograman	IoT	Data Challenge	Teknologi	Nasional	Individu	5
MHS029	Teknik Informatika	1	3.67	Keamanan Siber	Jaringan	AI	Hackathon	Teknologi	Internasional	Tim	4
MHS030	Teknik Informatika	5	3.29	Keamanan Siber	Basis Data	IoT	Data Challenge	Data	Nasional	Tim	5

Gambar 3. Contoh Dataset

4.2 PERANCANGAN MODEL

Proses generate model rekomendasi dilakukan dengan langkah-langkah berikut ini :

- Persiapan library. Pada penelitian ini beberapa library yang digunakan adalah pandas, numpy, tfidfvectorizer, math dan model_selection.

```

!pip install scikit-surprise

collecting scikit-surprise
Downloading scikit_surprise-1.1.4.tar.gz (154 kB)
154.4/154.4 kB 3.7 MB/s eta 0:00:00

Installing build dependencies ... done
Getting requirements to build wheel ... done
Preparing metadata (pyproject.toml) ... done
Requirement already satisfied: joblib>1.2.0 in /usr/local/lib/python3.10/dist-packages (from scikit-surprise) (1.4.2)
Requirement already satisfied: numpy>1.19.5 in /usr/local/lib/python3.10/dist-packages (from scikit-surprise) (1.26.4)
Requirement already satisfied: scipy>1.6.0 in /usr/local/lib/python3.10/dist-packages (from scikit-surprise) (1.13.1)
Building wheels for collected packages: scikit-surprise
Building wheel for scikit-surprise (pyproject.toml) ... done
Created wheel for scikit-surprise: filename=scikit_surprise-1.1.4-cp310-cp310-linux_x86_64.whl size=2357285 sha256=53e8f463ff1950863a948ac41e40896d7e45c66c96d4668769997ffe8d30bb3
Stored in directory: /root/.cache/pip/wheels/4b/3f/df/6acbf0a40397090f3ff97f5b2cc22f09ce66adde75bc71f054
Successfully built scikit-surprise
Installing collected packages: scikit-surprise
Successfully installed scikit-surprise-1.1.4

import pandas as pd
import numpy as np
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import linear_kernel
from sklearn.metrics import mean_squared_error
from math import sqrt
from surprise import Reader, Dataset, SVD
from surprise.model_selection import train_test_split
from surprise.accuracy import rmse

```

Gambar 4. Tangkapan layer proses persiapan Library

- Load data dari sumber format .csv dan memastikan data yang dimasukkan sesuai dengan sumbernya.

```

data_path = '/content/dataset_kompetensi_mahasiswa.csv'
df = pd.read_csv(data_path)
df.head()

mahasiswa_id      prodi  semester  ipk  kompetensi_1  kompetensi_2  minat  lomba  bidang_lomba  tingkat_lomba  jenis_lomba  rating_partisipasi
0      MHS001  Rekayasa Perangkat Lunak      4  3.60      Data Science      UI/UX  Networking  Debat      Desain      Lokal      Individu      5
1      MHS002      Teknik Komputer      8  3.65      Data Science  Keamanan Siber  Mobile App  Olimpiade IT      Desain      Lokal      Individu      2
2      MHS003      Informatika      6  3.65      Data Science      UI/UX  Data Analysis  Debat      Data      Internasional      Tim      5
3      MHS004      Sistem Informasi      6  2.61  Manajemen Proyek  Pemrograman      AI  Business Plan  Desain      Lokal      Individu      1
4      MHS005  Rekayasa Perangkat Lunak      5  3.76      Jaringan      Jaringan  Networking  Business Plan  Desain      Internasional      Individu      5

```

Gambar 5. Tangkapan Layar Proses Load Data

- Get recommendations based on content, pada tahapan ini dilakukan proses pencarian kesamaan rekomendasi lomba berdasarkan nilai preferensi dari pengguna. Implementasi algoritma *Item-Based Collaborative Filtering* untuk memberikan rekomendasi jenis lomba yang sesuai dengan kompetensi mahasiswa. Pada tahapan ini, data yang dihitung berupa pilihan jenis lomba dari preferensi mahasiswa lain yang diminati mahasiswa dan telah dikonversi ke dalam bentuk rating.

```
df['Keterampilan'] = df['Keterampilan'].fillna('').str.lower() # Lowercase and fill NAs
df['Keterampilan'] = df['Keterampilan'].str.replace('[^\w\s]', '', regex=True) # Remove punctuation

# Step 3: Content-Based Filtering using TF-IDF
tfidf = TfidfVectorizer(stop_words='english')
tfidf_matrix = tfidf.fit_transform(df['Keterampilan']) # Transform competency descriptions

# Calculate cosine similarity
cosine_sim = linear_kernel(tfidf_matrix, tfidf_matrix)

# Function to get recommendations based on content
def get_recommendations(Username, cosine_sim=cosine_sim):
    idx = df.index[df['Username'] == Username].tolist()[0] # Find index for student
    sim_scores = list(enumerate(cosine_sim[idx]))
    sim_scores = sorted(sim_scores, key=lambda x: x[1], reverse=True)
    sim_scores = sim_scores[1:11] # Get top 10 similar competencies
    student_indices = [i[0] for i in sim_scores]
    return df.iloc[student_indices]

# Test content-based recommendation
print(get_recommendations(Username='irwan21tet@mahasiswa.pcr.ac.id'))
```

	Timestamp	Username \
2	2024/07/15 7:35:44 PM GMT+7	taufik21tet@mahasiswa.pcr.ac.id
6	2024/07/15 7:41:36 PM GMT+7	lydia21tet@mahasiswa.pcr.ac.id
20	2024/10/17 6:48:13 PM GMT+7	sri22trk@mahasiswa.pcr.ac.id
23	2024/10/17 6:55:01 PM GMT+7	said23ti@mahasiswa.pcr.ac.id
26	2024/10/17 7:06:43 PM GMT+7	salwa23si@mahasiswa.pcr.ac.id
34	2024/10/17 8:29:04 PM GMT+7	christo21ti@mahasiswa.pcr.ac.id
35	2024/10/17 9:16:19 PM GMT+7	halimah23si@mahasiswa.pcr.ac.id
13	2024/07/16 11:27:09 PM GMT+7	rano21tet@mahasiswa.pcr.ac.id
18	2024/07/16 6:40:54 PM GMT+7	erdita22trk@mahasiswa.pcr.ac.id
22	2024/10/17 6:54:27 PM GMT+7	nur23trjt@mahasiswa.pcr.ac.id

	NamaLengkap	JenisKelamin \
2	Taufik Hidayat	Laki-laki
6	Lydia Nora Gisella S	Perempuan
20	Sri wahyuni	Perempuan
23	Tengku Said Arkan Raihan	Laki-laki
26	Salwa istiqfarah	Perempuan
34	Christo Ephesian Napitupulu	Laki-laki
35	Halimah Tussa'Diah	Perempuan
13	Rano putra sinaga	Laki-laki
18	Erdita Katri	Perempuan
22	Nur Anisa Handayani	Perempuan

Gambar 6. Get Recommendation

Terdapat 3 tahapan yang dilakukan pada penerapan algoritma *Item-Based Collaborative Filtering* yaitu data understanding yang merupakan tahap awal dalam memahami data yang akan digunakan, tahapan data preprocessing, dan tahapan ketiga adalah mendapatkan rekomendasi jenis lomba yang sesuai dengan preferensi mahasiswa lain [8].

- iv) Generate Model dengan *Naive Bayes*, Algoritma ini akan mengambil nilai IPK dari setiap mahasiswa sebagai input, dengan kata lain proses klasifikasi dilakukan untuk menentukan jenis lomba bagi mahasiswa berdasarkan IPK, kompetensi dan preferensi dari mahasiswa lain sebagai referensi. Terapat 4 tahapan dalam menjalankan algoritma ini yaitu proses pemahaman data (*data understanding*), *data preprocessing*, *modelling* dan klasifikasi jenis lomba.

```
nb = MultinomialNB()
nb.fit(X_train, y_train)

# Prediksi
y_pred = nb.predict(X_test)

# Evaluasi
print("Accuracy:", accuracy_score(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

Gambar 7. Proses Generate Model

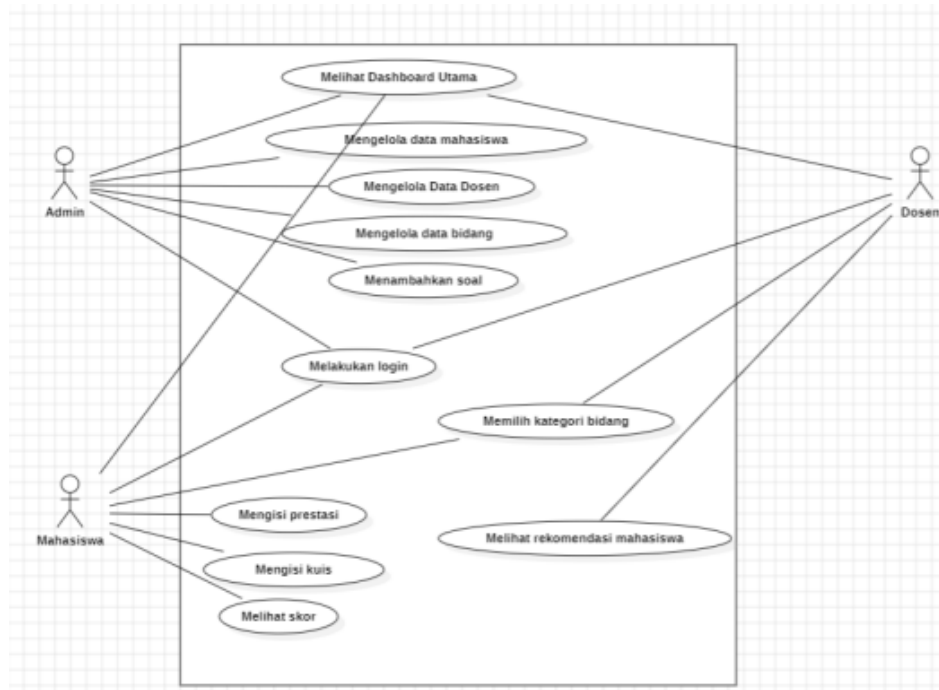
4.3 PERANCANGAN APLIKASI WEB

PowerMap adalah aplikasi berbasis web yang mendukung tema Pendidikan berkualitas di Indonesia, yang mana aplikasi ini menguji keterampilan dan kemampuan mahasiswa serta mengelola data-data prestasi mahasiswa. Sistem ini dapat diakses di lingkungan sistem operasi apapun yang dapat menjalankan web browser (Chrome, Edge, Firefox, Safari, dan lainnya). Ada beberapa fungsi yang bisa dilakukan oleh sistem ini, seperti melihat data mahasiswa berprestasi, mengelola data-data mahasiswa yang telah diuji, memberikan saran/rekomendasi siswa berprestasi yang dapat dilombakan ke ajang-ajang nasional maupun internasional. Berikut adalah karakteristik pengguna yang dikembangkan dalam aplikasi ini:

Tabel 1. Jenis Pengguna Sistem

No	Jenis Pengguna	Deskripsi
1	Operator	Aktor ini bertugas mengelola data katogori lomba, soal-soal kuis dan data mahasiswa
2	Dosen	Aktor ini bertugas untuk melihat daftar mahasiswa berprestasi dan data hasil pengujian atau rekomendasi jenis perlombaan
3	Mahasiswa	Aktorini bertugas untuk mengerjakan soal sebagai preferensi test dan merubah data diri

Dari hasil identifikasi data dan identifikasi kebutuhan pengguna, selanjutnya akan dilakukan analisis dan perancangan secara teknis. Beberapa data yang akan digunakan diantaranya data profil mahasiswa, data nilai akademik, data dan informasi lomba, serta data lainnya. Selanjutnya akan dilakukan EDA dan ETL untuk melihat karakteristik dari data. Proses pelabelan data akan dilakukan dengan menggunakan metode preferensi pengguna melalui survei dan wawancara dari data kompetensi mahasiswa dan keikutsertaan lomba. Selain itu pada tahapan ini juga akan dilakukan perancangan proses dengan menggunakan usecase diagram dan class diagram, serta perancangan wireframe untuk system yang akan dibuat.



Gambar 8. Usecase Diagram

Use case diagram untuk sistem rekomendasi siswa dapat digambarkan sebagaimana gambar di atas. Sistem dibangun dengan mengklasifikasi user menjadi 3 role akses, yaitu sebagai admin, yang bertanggung jawab untuk mengelola data mahasiswa dan dosen, mengelola bidang, dan mengelola hak akses untuk user. Disamping itu terdapat role mahasiswa yang memiliki fungsi mengisi Riwayat prestasi, mengisi kuis, serta selesai dikerjakan kuis akan menampilkan skor dan mahasiswa bisa melihatnya. Dan terakhir ada dosen sebagai user yang bisa melihat rekomendasi mahasiswa yang memiliki nilai akademis yang bagus serta nilai kuis yang tinggi berdasarkan bidang yang dipilihnya. Tahapan selanjutnya adalah melakukan implementasi terhadap perancangan yang telah dibuat.

5. PEMBAHASAN

Beberapa tahap yang akan dilakukan adalah membuat pemodelan data dengan naïve bayes dan item-base collaborative filtering, kemudian memberikan bobot dan melakukan perhitungan hybrid. Setelah itu model yang didapatkan akan diintegrasikan kedalam sebuah system berbasis web yang akan dibangun dengan menggunakan Bahasa pemrograman HTML, JavaScript dan PHP, serta dibantu dengan API Flask untuk mengintegrasikan model data rekomendasi yang sudah dibuat. Label data yang akan

digunakan pada proses klasifikasi data adalah jenis lomba yang terdiri dari : Pemrograman, Database, UI/UX, Jaringan, Game, Data Mining, dan Karya Tulis Ilmiah. Hasil dari proses ini adalah antarmuka system yang dibuat:



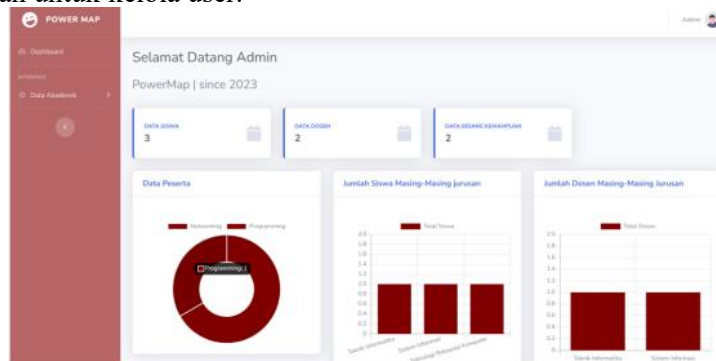
Gambar 9. Halaman Depan

Pada dashboard ini semua entitas yang ingin mengakses website ini terlebih dahulu akan diarahkan ke halaman ini, berisi informasi tentang aplikasi ini, dan terdapat tombol login. Halaman login adalah halaman yang diakses ketika user ingin mengakses sistem. Pada halaman ini dibutuhkan dua parameter yaitu email dan password.



Gambar 10. Halaman Login

Jika username dan password yang diinputkan terdapat di tabel admin, maka aplikasi akan mengarahkan ke dashboard admin yang berisi data-data user yang mengakses aplikasi dan mengerjakan soal lomba, serta terdapat inputan untuk kelola user.



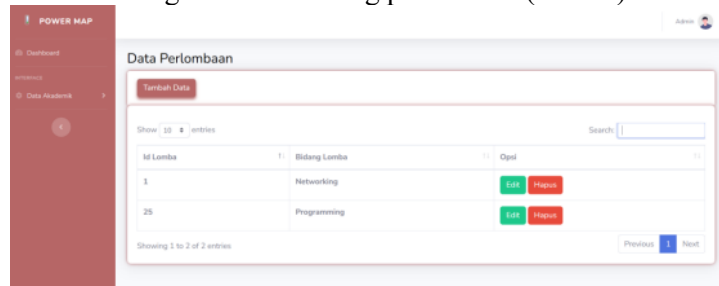
Gambar 11. Halaman Admin

Halaman ini dilakukan untuk mengelola data mahasiswa(CRUD)

No	Nim	Nama	Jurusan	Opai
1	22070409	satria	Sistem Informasi	<button>Edit</button> <button>Hapus</button>
2	22064321	Diana	Teknologi Rekayasa Komputer	<button>Edit</button> <button>Hapus</button>
3	220630	Aisyah Dwi	Teknik Informatika	<button>Edit</button> <button>Hapus</button>

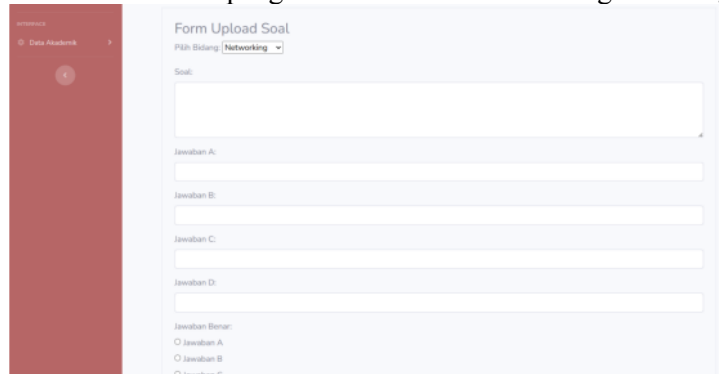
Gambar 12. Halaman Mahasiswa

Halaman ini dilakukan untuk mengelola data bidang perlombaan(CRUD)



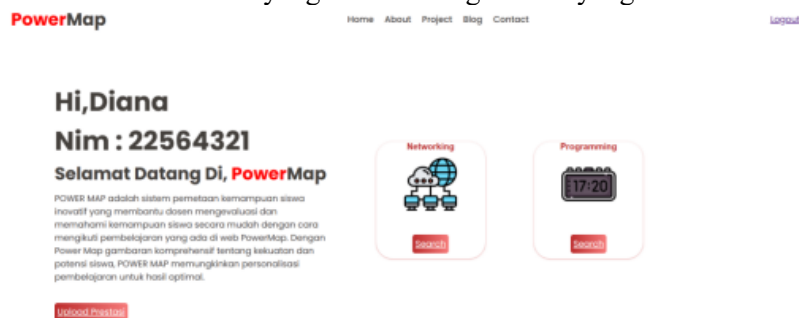
Gambar 13. CRUD Perlombaan

Pada halaman ini berisi formulir untuk pengisian soal berdasarkan kategori bidang.



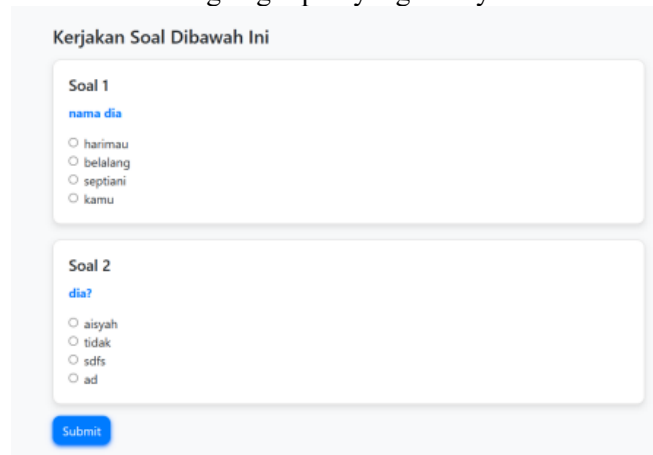
Gambar 14. Halaman Upload Soal

Jika username dan password yang diinputkan terdapat di tabel mahasiswa,maka aplikasi akan mengarahkan ke dashboard mahasiswa yang berisi bidang/matkul yang akan dilombakan.



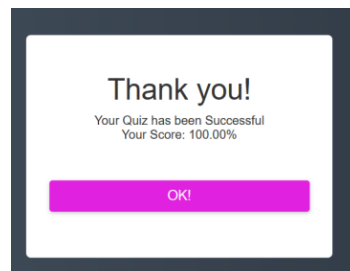
Gambar 15. Halaman Mahasiswa

Merupakan lembar soal dari setiap bidang yang dipilih, jika mahasiswa memilih bidang networking maka akan menampilkan soal networking begitupun yang lainnya.



Gambar 16. Halaman Soal

Wireframe ini muncul setelah mahasiswa selesai upload jawaban,dan dihitung total benar dari seluruh jumlah soal yg ada.

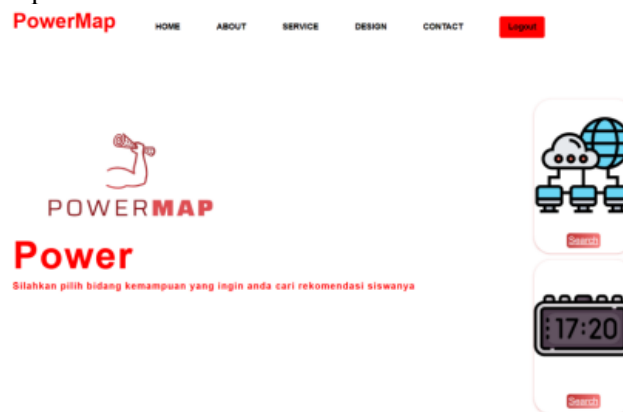


Gambar 17. Pop Up Skor

Halaman ini merupakan menu inputan prestasi bagi mahasiswa,didukung oleh sertifikat prestasi

Gambar 18. Upload Prestasi

Jika username dan password yang diinputkan terdapat di tabel dosen, maka aplikasi akan mengarahkan ke dashboard dosen yang terdapat list menu bidang rekomendasi mahasiswa, lalu setelah dosen memilih bidang tersebut akan ditampilkan list-list rekomendasi mahasiswa.



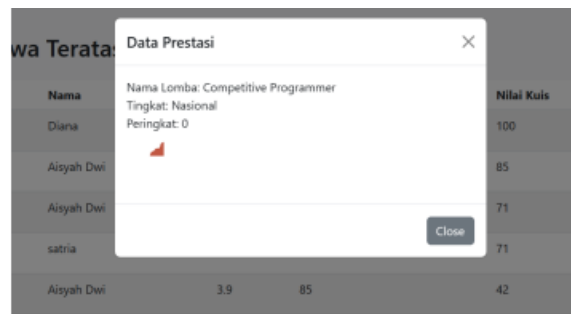
Gambar 19. Halaman Dosen

Merupakan halaman rekomendasi yang menampilkan siswa unggulan di bidang yang ingin dosen cari mahasiswa unggulnya

No	NIM	Nama	IPK	Nilai Bidang	Nilai Kuis	Prestasi
1	22564321	Diana	3.98	70	100	
2	225630	Aisyah Dai	3.9	85	85	
3	225630	Aisyah Dai	3.9	85	71	
4	22676489	satira	0	0	71	
5	225630	Aisyah Dai	3.9	85	42	
6	225630	Aisyah Dai	3.9	85	42	
7	225630	Aisyah Dai	3.9	85	42	
8	225630	Aisyah Dai	3.9	85	42	
9	225630	Aisyah Dai	3.9	85	28	

Gambar 20. Halaman Rekomendasi

Pop Up ini muncul ketika dosen melihat rekomendasi mahasiswa serta melakukan klik button pada kolom prestasi.



Gambar 21. Pop Up Prestasi

Penggunaan hybrid recommendation pada penelitian ini yaitu *Naïve Bayes* dan *Item-Based Collaborative Filtering* dengan kontribusi dari masing-masing metode adalah sebagai berikut : Algoritma *Naïve Bayes* digunakan untuk klasifikasi berdasarkan atribut seperti prodi, semester, ipk, kompetensi, dan minat. Algoritma ini memiliki kemampuan model dalam mempelajari distribusi sederhana antar fitur membantu dalam generalisasi awal. Pelabelan yang dilakukan oleh *Naïve Bayes* menghasilkan rekomendasi jenis lomba berdasarkan kemiripan historis antar lomba (*bidang_lomba*, *jenis_lomba*, *tingkat_lomba*). Sementara kontribusi *Item-Based collaborative* memperkuat hasil prediksi dari *Naïve Bayes* dengan mempertimbangkan preferensi mahasiswa serupa. Cara kerja dari pendekatan ini adalah sebagai berikut :

- i) Membuat matriks yang merepresentasikan interaksi antara pengguna dalam hal ini mahasiswa dengan atribut lainnya yaitu jenis lomba. Matriks ini berisi skor preferensi atau rating
- ii) Menghitung kemiripan antar item dengan menggunakan matrik cosine similarity atau pearson correlation
- iii) Selanjutnya membangun model kemiripan serta melakukan prediksi dan rekomendasi.

5.1 EVALUASI PENGUJIAN SISTEM

Hasil evaluasi dari model yang telah dilakukan mampu memprediksi dengan benar sebesar 70% adalah:

Accuracy: 0.7					
	precision	recall	f1-score	support	
0	0.75	0.75	0.75	24	
1	0.62	0.62	0.62	16	
accuracy			0.70	40	
macro avg	0.69	0.69	0.69	40	
weighted avg	0.70	0.70	0.70	40	

Gambar 22. Akurasi Model

Model rekomendasi yang dibangun dalam penelitian ini menunjukkan performa yang cukup baik, dengan tingkat akurasi mencapai 70% dalam memprediksi jenis lomba yang sesuai bagi mahasiswa. Persentase ini mengindikasikan bahwa sistem mampu memberikan rekomendasi yang relevan dalam sebagian besar kasus. Capaian ini mencerminkan keberhasilan model dalam mengenali pola hubungan antara data akademik dan minat mahasiswa terhadap partisipasi dalam kompetisi.

Tabel 2. Perbandingan Akurasi dengan Metode Lain

Metode	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)
Naïve Bayes	70	75	75	75
Random Forest	70	50	46	48
SVM	69	48	46	47

Berdasarkan hasil perbandingan akurasi yang dilakukan antara metode naïve bayes dengan random forest dan SVM didapatkan hasil akurasi paling tinggi adalah naïve bayes, hal ini dipengaruhi oleh penggunaan label encoding yang menghasilkan fitur numerik yang cocok dengan karakteri naïve bayes. Selain itu berdasarkan penelitian yang dilakukan oleh [19] menyatakan bahwa naïve bayes relative lebih

tahan terhadap overfitting di dataset yang kecil, sehingga pada penelitian ini dengan dataset yang kecil performanya relatif lebih stabil.

Keberhasilan model tidak terlepas dari kontribusi dua pendekatan yang digunakan, yaitu Naïve Bayes dan *Item-Based Collaborative Filtering*. Naïve Bayes berperan sebagai algoritma klasifikasi utama yang memanfaatkan data atribut mahasiswa, seperti program studi, semester, IPK, kompetensi, dan minat. Sifat dasar algoritma ini yang mengasumsikan independensi antar fitur memungkinkan proses klasifikasi dilakukan dengan efisien.

Di sisi lain, metode *Item-Based Collaborative Filtering* berkontribusi dalam memperkuat rekomendasi dengan menambahkan informasi berbasis kesamaan perilaku antar mahasiswa. Metode ini mengevaluasi kemiripan antar jenis lomba berdasarkan pola partisipasi sebelumnya dan mengidentifikasi kecenderungan mahasiswa yang memiliki profil mirip untuk mengikuti lomba tertentu. Dengan mempertimbangkan konteks kolektif dari preferensi mahasiswa, metode ini melengkapi pendekatan berbasis atribut yang digunakan oleh Naïve Bayes.

Integrasi kedua metode tersebut menghasilkan pendekatan yang lebih holistik dalam membangun sistem rekomendasi. Naïve Bayes memberikan prediksi berdasarkan karakteristik individu, sedangkan Collaborative Filtering menambahkan nilai dari sisi kesamaan perilaku antar pengguna. Sinergi ini berperan dalam meningkatkan ketepatan dan relevansi hasil rekomendasi, sehingga sistem dapat berkontribusi secara lebih efektif dalam mendukung pengembangan potensi mahasiswa melalui pemilihan lomba yang sesuai.

6. KESIMPULAN

Penelitian ini bertujuan mengembangkan sistem rekomendasi berbasis hybrid recommendation untuk memetakan kompetensi mahasiswa terhadap jenis lomba yang relevan. Model yang dikembangkan menggunakan kombinasi metode content-based filtering berbasis TF-IDF dan cosine similarity, serta algoritma Naïve Bayes untuk prediksi. Evaluasi menunjukkan akurasi model sebesar 70%, yang mencerminkan kemampuan model untuk memberikan rekomendasi, meskipun masih ada ruang untuk perbaikan.

Secara umum, hasil penelitian ini menunjukkan bahwa pendekatan hybrid recommendation dapat diterapkan untuk membantu proses pemetaan kompetensi mahasiswa terhadap lomba, tetapi diperlukan pengayaan data dan optimalisasi metode untuk meningkatkan akurasi dan relevansi rekomendasi. Penelitian ini memberikan kontribusi penting dalam bidang sistem rekomendasi, khususnya untuk aplikasi pendidikan, dengan potensi pengembangan lebih lanjut untuk hasil yang lebih optimal.

Berdasarkan hasil penelitian yang telah dilakukan kemungkinan adanya asumsi independensi fitur dalam *Naive Bayes* yang tidak selalu mencerminkan kondisi sebenarnya, khususnya pada dataset penelitian ini ketika minat dan kompetensi saling berkaitan. Hal ini dapat menyebabkan overgeneralisasi atau salah klasifikasi ketika dataset yang diolah terdapat data kasus lintas prodi. Sehingga penelitian selanjutnya dapat menggunakan metode lain yang lebih baik tanpa terpengaruh dari independensi fitur yang digunakan.

UCAPAN TERIMA KASIH

Terima kasih kami ucapkan kepada Politeknik Caltex Riau yang secara teknis dikelola oleh BP2M atas hibah internal yang telah diberikan dalam mendukung penelitian ini.

DAFTAR PUSTAKA

- [1] R. Setiamihardja, "Portofolio Assessment," *J. Pendidik. Dasar Kampus Cibiru*, vol. 3, no. 2, pp. 1–2, 2012, [Online]. Available: <https://ejournal.upi.edu/index.php/eduhumaniora/article/view/2806/1832>
- [2] N. Asmilia, "Pentingnya Portofolio Investasi untuk Mengetahui Keanekaragaman Resiko dan Meningkatkan Minat Berinvestasi di Kalangan Generasi Milenial," *J. Pengabd. Kpd. Masyarakat*, vol. 4, no. 3, 2023.
- [3] M. M. Dewi, R. Andriani, and M. Nuraminudin, "Performance Analysis of the *Item-Based Collaborative Filtering* Model in Yogyakarta Tourism Recommendations," vol. 9, no. 2, pp. 534–541, 2025.

- [4] J. Aisyiah and L. Cahyani, "Sistem Rekomendasi Program Studi Menggunakan Metode Hybrid Recommendation (Studi Kasus: MAN Sumenep)," pp. 59–72, 2023, doi: 10.30864/eksplora.v12i1.992.
- [5] D. S. Pradana, P. Prajoko, and G. P. Hartawan, "Perbandingan Algoritma Content-Based Filtering dan Collaborative Filtering dalam Rekomendasi Kegiatan Ekstrakurikuler Siswa," *Progresif J. Ilm. Komput.*, vol. 18, no. 2, p. 151, 2022, doi: 10.35889/progresif.v18i2.854.
- [6] M. I. Rizky, I. Asror, and Y. R. Murti, "Sistem Rekomendasi Program Studi untuk Siswa SMA Sederajat Menggunakan Metode Hybrid Recommendation dengan Content Based Filtering dan Collaborative Filtering," in *e-Proceeding of Engineering*, 2020, p. 17.
- [7] L. Tommy, D. Novianto, and Y. S. Japriadi, "Sistem Rekomendasi Hybrid untuk Pemesanan Hidangan Berdasarkan Karakteristik dan Rating Hidangan," *J. Appl. Informatics Comput.*, vol. 4, no. 2, pp. 137–145, 2020, doi: 10.30871/jaic.v4i2.2687.
- [8] A. Yusmar, *Implementasi metode collaborative filtering dengan pendekatan Item-Based untuk rekomendasi rumah makan menggunakan algoritma adjusted cosine similarity*. 2020. [Online]. Available: [https://repository.uinjkt.ac.id/dspace/handle/123456789/55992%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/55992/1/ADDINI YUSMAR-FST.pdf](https://repository.uinjkt.ac.id/dspace/handle/123456789/55992%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/55992/1/ADDINI%20YUSMAR-FST.pdf)
- [9] A. A. Widjaja and H. N. Palit, "Hybrid Recommendation System untuk Peminjaman Buku Perpustakaan dengan Collaborative dan Content-Based Filtering," *J. Infra*, vol. 10, no. 2, pp. 1–6, 2022, [Online]. Available: <https://publication.petra.ac.id/index.php/teknik-informatika/article/view/12512>
- [10] M. Nurizki, W. Apriandari, and A. Asriyanik, "Algoritma Naïve Bayes untuk Rekomendasi Seleksi Peserta Paskibraka Berbasis Website," *J. Inf. Syst. Res.*, vol. 4, no. 4, pp. 1486–1493, 2023, doi: 10.47065/josh.v4i4.3574.
- [11] Y. Dananjoyo and Y. Muflilah, "Sistem Rekomendasi Ukuran Baju Pada Aplikasi E-Commece Dengan Metode Naïve Bayes Yogo," *Jasmatika*, vol. 3, no. 1, pp. 10–15, 2015.
- [12] R. Sudhir, "A Survey on Image Mining Techniques : Theory and Applications," vol. 2, no. 6, pp. 44–53, 2011.
- [13] S. Easterbrook and J. Callahan, "Formal Methods for Verification and Validation of partial specifications : A Case Study 1 Introduction 2 Context : The IV & V Process," pp. 1–13.
- [14] V. Srividhya and R. Anitha, "Evaluating Preprocessing Techniques in Text Categor ization," pp. 49–51, 2010.
- [15] N. Aini *et al.*, "Implementasi Metode Hybrid Recommendation untuk Sistem Rekomendasi Mitra MSIB di Prodi Pendidikan Informatika," vol. 10, no. 2, pp. 171–180, 2024.
- [16] L. Zhu and B. W. B., *on Hybrid Recommendation Algorithms*. Atlantis Press International BV. doi: 10.2991/978-94-6463-044-2.
- [17] Z. Zhang, T. Peng, and K. Shen, "Overview of Collaborative Filtering Recommendation Algorithms," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 440, no. 2, 2020, doi: 10.1088/1755-1315/440/2/022063.
- [18] E. G. F. H. del Olmo, "Evaluation of recommender systems: A new approach," *Expert Syst. Appl.*, pp. 1–15, 2008.
- [19] V. Nakhipova *et al.*, "Use of the Naive Bayes Classifier Algorithm in Machine Learning for Student Performance Prediction," *Int. J. Inf. Educ. Technol.*, vol. 14, no. 1, pp. 92–98, 2024, doi: 10.18178/ijiet.2024.14.1.2028.