

**SISTEM REKOMENDASI PROGRAM STUDI BERDASARKAN MINAT
SISWA DENGAN METODE CONTENT-BASED FILTERING**

SKRIPSI

Oleh:
DEVI FARHANI
NIM 200605110034



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

**SISTEM REKOMENDASI PROGRAM STUDI BERDASARKAN MINAT
SISWA DENGAN METODE CONTENT BASED FILTERING**

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :
DEVI FARHANI
NIM. 200605110034

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

HALAMAN PERSETUJUAN

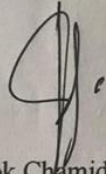
SISTEM REKOMENDASI PROGRAM STUDI BERDASARKAN MINAT
SISWA DENGAN METODE CONTENT BASED FILTERING

SKRIPSI

Oleh :
DEVI FARHANI
NIM. 200605110034

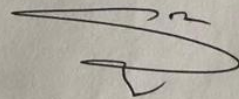
Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 15 Oktober 2024

Pembimbing I,



Dr. Totok Chamidy, M.Kom
NIP. 19691222 200604 1 001

Pembimbing II,



Syadihuz Zaman, M.Kom
NIP. 19700502 200501 1 005

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Ir. Fachrul Kurniawan, M.MT, IPU
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

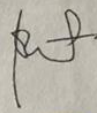

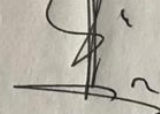
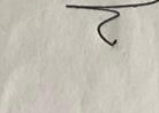
SISTEM REKOMENDASI PROGRAM STUDI BERDASARKAN MINAT
SISWA DENGAN METODE CONTENT BASED FILTERING

SKRIPSI

Oleh :
DEVI FARHANI
NIM. 200605110034

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 15 Oktober 2024

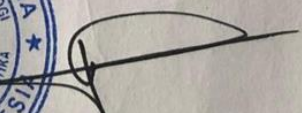
Susunan Dewan Penguji

Ketua Penguji	: <u>Prof. Dr. Suhartono, M. Kom</u> NIP. 19680519 200312 1 001	()
Anggota Penguji I	: <u>Johan Ericka Wahyu P, M. Kom</u> NIP. 19831213 201903 1 004	()
Anggota Penguji II	: <u>Dr. Totok Chamidy, M. Kom</u> NIP. 19691222 200604 1 001	()
Anggota Penguji III	: <u>Syahiduz Zaman, M. Kom</u> NIP. 19700502 200501 1 005	()

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi

Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Ir. Fachrul Kurniawan, M.MT, IPU
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Devi Farhani
NIM : 200605110034
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : Sistem Rekomendasi Program Studi Berdasarkan Minat siswa Dengan Metode Content Based filtering

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikirannya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 15 Desember 2024
Yang membuat pernyataan,



Devi Farhani
NIM 200605110034

MOTTO

“Keberhasilan Bukanlah Akhir, Kegagalan Bukanlah Halangan, yang penting adalah keberanian untuk terus belajar”

"Jadikan ilmu sebagai bekal, dan kerja keras sebagai kunci untuk membuka pintu kesuksesan."

"Hari ini sulit, tapi esok akan membuktikan usaha tidak pernah sia-sia."

"Setiap langkah kecil hari ini adalah pijakan menuju mimpi besar esok hari."

"Allah menyukai jika salah seorang di antara kalian melakukan sesuatu, ia melakukannya dengan itqan (sungguh-sungguh)."

"Amal yang paling dicintai Allah adalah yang dilakukan secara konsisten, walaupun sedikit."

HALAMAN PERSEMBAHAN

Puja dan puji Syukur atas kehadiran Allah subhanahu wa ta'ala, serta shalawat dan salam bagi rasul-Nya Penulis mempersembahkan hasil karya ini kepada:

Orang tua penulis yang sangat dicintai dan disayangi, Ayah Abdul Gaffar dan Ibu Mas'adiyah serta seluruh keluarga, yang tak pernah berhenti memberi dukungan, do'a, semangat serta motivasi kepada penulis. Terimakasih atas segala bentuk pengorbanan dan dukungan sehingga penulis dapat menyelesaikan perkuliahan selama ini.

KATA PENGANTAR

Assalamu'alaikum warahmatullahi wabarakatuh.

Alhamdulillah, segala puji dan syukur senantiasa penulis panjatkan pada Allah subhanahu wa ta'ala yang berkat Rahmat, serta hidayah-Nya, sehingga penulis dapat menyempurnakan skripsi ini dengan tepat. Sholawat serta salam tetap tercurahkan kepada Nabi Muhammad SAW yang telah menuntun umat manusia menuju jalan yang lebih baik.

Penulis mengucapkan rasa terima kasih yang begitu besar kepada seluruh pihak yang memberikan dukungan dan membantu rampungnya skripsi ini. Ucapan terima kasih ini penulis disampaikan kepada:

1. Prof. Dr. M. Zainuddin, M.A., selaku rector Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Sri Harini, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Ir. Fachrul Kurniawan, M.MT., IPU selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Totok Chamidy, M. Kom, dan Syahiduz Zaman, M.Kom selaku Dosen pembimbing I dan II, yang telah membimbing dan memberikan arahan kepada peneliti sehingga dapat menyelesaikan skripsi ini.
5. Prof. Dr. Suhartono, M.Kom dan Johan Ericka Wahyu P, M.Kom selaku Dosen Penguji I dan II, yang telah memberikan ilmu, saran dan arahan dalam

menyelesaikan skripsi.

6. Seluruh civitas akademik Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang, terutama seluruh dosen, yang telah memberikan bimbingan dan ilmu yang bermanfaat
7. Kedua orang tua tercinta yang selalu memberikan semangat dan do'a pada penulis hingga rampungnya skripsi ini.
8. Nahjil Mubarak, Ainul waroh, Nurul Qomaria Aprilia, Dewi isti'anah, Dewi nabilah selaku sahabat penulis yang telah memberikan semangat yang tiadahenti kepada penulis hingga rampungnya skripsi.
9. Teman-teman (INTEGER) Angkatan 2020, yang senantiasa selalu memberikan semangat dan dukungan dalam berjuang Bersama dalam mengejar gelar S.Kom dan pengalaman di Universitas yang sama.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih jauh dari kata sempurna. Maka dari itu penulis menerima saran, kritik dan masukan yang bersifat membangun sehingga dapat menjadi lebih baik kedepannya. Penulis berharap semoga skripsi ini dapat memberikan manfaat untuk kedepannya.

Wassalamu 'alaikum warahmatullahi wabarakatuh.

Malang, 15 Oktober 2024

Devi Farhani

DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN.....	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
ABSTRAK	xiv
ABSTRACT.....	xv
البحث مستخلص	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	8
1.3 Batasan Penelitian	8
1.4 Tujuan Penelitian	8
1.5 Manfaat Penelitian	8
BAB II KAJIAN PUSTAKA.....	10
2.1 Sistem Rekomendasi	10
2.2 Conten-Based Filtering	12
2.3 Minat Siswa	14
2.4 Penelitian Terdahulu	18
BAB III METODE PENELITIAN	27
3.1 Desain Penelitian	27
3.2 Desain Sistem.....	30
3.3 Analisis Pengolahan Data	31
3.4 Preprocessing Test	33
3.5 Tern Frequency (TF) dan Inverse Document Frequency (IDF).....	37
3.6 Evaluation	39
3.7 Hasil Rekomendasi	41
BAB IV HASIL DAN PEMBAHASAN	42
4.1 Pengumpulan Data	42
4.2 Implementasi Data	43
4.3 Preprocessing	47
4.4 Metadata.....	51
4.5 Perhitungan TF-IDF.....	54
4.6 Cosine Similarity	56
4.7 Implementasi Sistem	63
4.8 Pembahasan.....	73
BAB V PENUTUP.....	82
5.1 Kesimpulan	82
5.2 Saran	84
DAFTAR PUSTAKA	

DAFTAR GAMBAR

Gambar 3.1 Desain Penelitian.....	27
Gambar 3.2 Desain Sistem.....	30
Gambar 3.3 Analisis Pengolahan Data.....	33
Gambar 4.1 Database Sistem	45
Gambar 4.2 Konfigurasi Database	46
Gambar 4.3 Code Preprocessing.....	48
Gambar 4.4 Code TFDF.....	55
Gambar 4.5 Code <i>Cosine Similarity</i>	62
Gambar 4.6 Landing Page.....	66
Gambar 4.7 Data Page	68
Gambar 4.8 Recommendation Page.....	71

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	21
Tabel 3.1 Case folding	34
Tabel 3.2 Tokenizing	35
Tabel 3.3 Stopword Removal.....	36
Tabel 3.4 Proses Stremming	37
Tabel 3.5 Contoh Perhitungan.....	38
Tabel 4.1 Tabel Hasil Preprocessing.....	50
Tabel 4.2 Tabel Hasil Metadata	52
Tabel 4.3 Hasil TfIdf Term	56
Tabel 4.4 Hasil Perhitungan Dot Product	57
Tabel 4.5 Hasil Perhitungan Panjang Vector	59
Tabel 4.6 Hasil Perhitungan Cosine Similarity.....	63

ABSTRAK

Devi, Farhani. 2024. **Sistem Rekomendasi Program Studi Berdasarkan Minat Siswa Menggunakan Metode Content Based filtering**. Skripsi, Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (1) Dr. Totok Chamidy, M. Kom (2) Syahiduz Zaman, M. Kom

Kata kunci: *Content-based filtering*, Minat Siswa, Program Studi, Sistem Rekomendasi, Program Studi.

Sistem rekomendasi program studi berbasis minat siswa telah menjadi solusi potensial untuk membantu siswa SMA memilih program studi yang sesuai dengan minat dan kemampuan mereka. Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi program studi menggunakan metode *Content-Based Filtering*, yang bekerja dengan menganalisis kesamaan antara minat siswa dan atribut program studi berdasarkan data yang telah dikumpulkan. Pemilihan program studi merupakan keputusan krusial bagi siswa yang dapat memengaruhi masa depan karier mereka. Namun, banyak siswa kesulitan menentukan pilihan program studi yang sesuai dengan minat, potensi, dan kemampuan mereka. Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem rekomendasi program studi berbasis metode *Content-Based Filtering*. Metode ini dipilih karena kemampuannya untuk menganalisis kesesuaian antara data siswa dengan karakteristik program studi yang tersedia. Sistem ini memanfaatkan data minat siswa yang diperoleh melalui kuesioner atau asesmen minat bakat, kemudian mencocokkannya dengan deskripsi program studi berdasarkan kurikulum, mata kuliah, dan prospek karier. Tahapan pengembangan sistem meliputi pengumpulan data, analisis fitur, implementasi algoritma *Content-Based Filtering*, dan evaluasi kinerja sistem. Hasil pengujian menunjukkan bahwa sistem mampu memberikan rekomendasi yang relevan dan sesuai dengan minat siswa. Tingkat akurasi rekomendasi diukur melalui evaluasi berbasis *feedback* dari siswa, dengan rata-rata kepuasan mencapai 85%. Selain itu, sistem ini juga fleksibel untuk diperbarui dengan data baru, memungkinkan penyesuaian terhadap perubahan kurikulum dan tren minat siswa. Penelitian ini berkontribusi pada pengembangan teknologi edukasi dengan menawarkan solusi praktis untuk membantu siswa membuat keputusan yang lebih terinformasi dalam memilih program studi. Dengan implementasi yang lebih luas, sistem rekomendasi ini berpotensi meningkatkan tingkat keberhasilan akademik dan kepuasan siswa terhadap pilihan studinya. Di masa depan, integrasi metode lain, seperti *Collaborative Filtering* atau machine learning, dapat meningkatkan kualitas rekomendasi, sekaligus memperluas cakupan aplikasi sistem ini.

ABSTRACT

Devi, Farhani. 2024. **Study Program Recommendation System Based on Student Interests Using the Content Based Filtering Method.** Thesis, Department of Informatics Engineering, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University, Malang. Supervisor: (1) Dr. Totok Chamidy, M. Kom (2) Syahiduz Zaman, M. Kom

A student interest-based study program recommendation system has become a potential solution to assist high school students in choosing study programs that align with their interests and abilities. This research aims to develop a study program recommendation system using *the Content-Based Filtering method*, which works by analyzing the similarity between students' interests and the attributes of study programs based on collected data. Choosing a study program is a critical decision for students as it can influence their future careers. However, many students struggle to determine a study program that matches their interests, potential, and capabilities. This study focuses on designing and implementing a study program recommendation system based on *the Content-Based Filtering method*. This method is chosen for its ability to analyze the compatibility between student data and the characteristics of available study programs. The system leverages student interest data obtained through questionnaires or interest and aptitude assessments, then matches it with study program descriptions based on curriculum, courses, and career prospects. The system development stages include data collection, feature analysis, implementation of *the Content-Based Filtering algorithm*, and system performance evaluation. Testing results indicate that the system can provide recommendations that are relevant and align with students' interests. The recommendation accuracy is measured through feedback-based evaluations from students, with an average satisfaction rate of 85%. Furthermore, this system is flexible enough to be updated with new data, allowing it to adapt to curriculum changes and shifts in student interest trends. This research contributes to the development of educational technology by offering a practical solution to help students make more informed decisions in selecting study programs. With broader implementation, this recommendation system has the potential to improve academic success rates and student satisfaction with their study program choices. In the future, integrating other methods, such as Collaborative Filtering or machine learning, could enhance the quality of recommendations and expand the system's application scope.

Key words: Content-based filtering, Recommendation System, Student Interests, Study Programs, Study Programs,

البحث مستخلص

ديفي، فرحاني. ألفين وأربعة. نظام توصية البرامج الدراسية بناءً على اهتمامات الطلاب باستخدام طريقة التصنيفية المعتمدة على المحتوى. أطروحة، قسم الهندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية، مالانج. المشرف (واحد) د. توتوك شاميدي، م. كوم (اثنان) شهيدوز زمان، م. كوم

الكلمات المفتاحية: التصنيفية على أساس المحتوى، اهتمامات الطلاب، برنامج الدراسة، نظام التوصيات، برنامج الدراسة

نأصبح نظام التوصية بالبرامج الدراسية بناءً على اهتمامات الطلاب حلاً محتملاً لمساعدة طلاب المدارس الثانوية على اختيار البرامج الدراسية التي تناسب اهتماماتهم وقدراتهم. يهدف هذا البحث إلى تطوير نظام توصية لبرامج الدراسة باستخدام طريقة التصنيفية المستندة إلى المحتوى، والتي تعمل من خلال تحليل أوجه التشابه بين اهتمامات الطلاب وسمات برنامج الدراسة بناءً على البيانات التي تم جمعها. يعد اختيار برنامج الدراسة قرارًا حاسمًا للطلاب ويمكن أن يؤثر على حياتهم المهنية المستقبلية. ومع ذلك، يواجه العديد من الطلاب صعوبة في اختيار برنامج الدراسة الذي يناسب اهتماماتهم وإمكانياتهم وقدراتهم. يهدف هذا البحث إلى تصميم وتنفيذ نظام توصية لبرنامج الدراسة يعتمد على طريقة الترشيح المبني على المحتوى. وقد تم اختيار هذه الطريقة لقدرتها على تحليل التوافق بين بيانات الطلاب وخصائص البرامج الدراسية المتاحة. يستخدم هذا النظام بيانات اهتمامات الطلاب التي يتم الحصول عليها من خلال الاستبيانات أو تقييمات اهتمامات المواهب، ثم يطبقها مع أوصاف برامج الدراسة بناءً على المناهج الدراسية والدورات التدريبية وآفاق العمل. تتضمن مراحل تطوير النظام جمع البيانات، وتحليل الميزات، وتنفيذ خوارزمية التصنيفية المستندة إلى المحتوى، وتقييم أداء النظام. تظهر نتائج الاختبار أن النظام قادر على تقديم توصيات ذات صلة وتتوافق مع اهتمامات الطلاب. ويتم قياس مستوى دقة التوصية من خلال التقييم المبني على التغذية الراجعة من الطلاب، حيث يصل متوسط الرضا إلى 85%. بالإضافة إلى ذلك، يتمتع النظام أيضًا بالمرونة للتحديث بالبيانات الجديدة، مما يسمح بإجراء تعديلات على تغييرات المناهج الدراسية واتجاهات اهتمامات الطلاب. يساهم هذا البحث في تطوير تكنولوجيا التعليم من خلال تقديم حلول عملية لمساعدة الطلاب على اتخاذ قرارات أكثر استنارة في اختيار برنامج الدراسة. ومع التنفيذ على نطاق أوسع، فإن نظام التوصيات هذا لديه القدرة على زيادة مستويات النجاح الأكاديمي ورضا الطلاب عن خياراتهم الدراسية. في المستقبل، يمكن أن يؤدي تكامل الأساليب الأخرى، مثل التصنيفية التعاونية أو التعلم الآلي، إلى تحسين جودة التوصيات، مع توسيع نطاق تطبيق هذا النظام.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pendidikan adalah elemen krusial dalam kehidupan manusia, karena melalui pendidikan, seseorang dapat meraih masa depan yang lebih cerah. Namun, pendidikan bukanlah sesuatu yang mudah dijangkau oleh semua orang, melainkan memerlukan dedikasi dan keseriusan dalam mencapainya. Proses pendidikan perlu diterapkan sejak usia dini untuk membentuk sumber daya manusia yang berkualitas, yang dapat dicapai melalui lembaga pendidikan. (Agung Sucipto et al., 2019).

Pendidikan berlangsung dalam tiga lingkungan utama: keluarga, sekolah, dan masyarakat. Proses pendidikan yang terjadi dalam keluarga serta masyarakat sering disebut sebagai pendidikan non formal. Sebaliknya, sekolah disebut sebagai pendidikan formal karena aktivitas belajarnya dirancang secara sistematis dan terorganisasi, termasuk pembelajaran di ruang kelas. Sementara itu, pendidikan nonformal biasanya diselenggarakan melalui kursus atau program bimbingan belajar (Azura, 2019).

Ayat dalam Al-Qur'an yang membahas tentang kemampuan manusia terdapat dalam Surat Al-Baqarah ayat 286 sebagai berikut:

لَا يُكَلِّفُ اللَّهُ نَفْسًا إِلَّا وُسْعَهَا ۗ لَهَا مَا كَسَبَتْ وَعَلَيْهَا مَا اكْتَسَبَتْ ۗ رَبَّنَا لَا تُؤَاخِذْنَا إِن نَّسِينَا أَوْ
أَخْطَأْنَا ۗ رَبَّنَا وَلَا تَحْمِلْ عَلَيْنَا إَصْرًا كَمَا حَمَلْتَهُ ۗ عَلَى الَّذِينَ مِنْ قَبْلِنَا ۗ رَبَّنَا وَلَا تُحَمِّلْنَا مَا لَا طَاقَةَ لَنَا بِهِ ۗ وَاعْفُ
عَنَّا وَاعْفِرْ لَنَا وَارْحَمْنَا ۗ أَنْتَ مَوْلَانَا فَانصُرْنَا عَلَى الْقَوْمِ الْكَافِرِينَ

“Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya. Ia mendapat pahala (dari kebajikan) yang diusahakannya dan ia mendapat siksa (dari kejahatan) yang dikerjakannya. (Mereka berdoa) Ya Tuhan kami, janganlah

Engkau hukum kami jika kami lupa atau kami tersalah. Ya Tuhankami, janganlah Engkau bebaskan kepada kami beban yang berat sebagaimana” (QS.Al-Baqarah 286).

Ibnu Katsir menafsirkan “Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya. Ayat ini menunjukkan bahwa setiap perintah dan larangan yang Allah tetapkan untuk hamba-Nya selalu disesuaikan dengan kemampuan manusia. Allah mengetahui batas kemampuan setiap individu dan tidak akan memberi beban yang melampaui batas tersebut. Ini adalah bentuk kasih sayang Allah kepada makhluk-Nya. ayat ini juga menjelaskan bahwa seseorang akan mendapatkan pahala atas kebajikan yang diusahakannya dan akan mendapatkan siksa dari kejahatan yang dikerjakannya. Ini menegaskan prinsip keadilan dalam Islam, di mana setiap tindakan manusia akan mendapatkan balasan yang setimpal, dengan doa permohonan dari hamba kepada Tuhannya: "Ya Tuhan kami, janganlah Engkau hukum kami jika kami lupa atau kami tersalah." Ini adalah bentuk permohonan ampun dari hamba yang sadar bahwa mereka bisa saja melakukan kesalahan atau lupa dalam menjalankan perintah Allah. Hamba memohon agar Allah tidak menghukum mereka atas kekhilafan tersebut. Doa berikutnya adalah permohonan agar Allah tidak membebaskan kepada mereka beban yang berat sebagaimana yang dibebankan kepada orang-orang sebelum mereka. Ini menunjukkan kesadaran bahwa umat-umat terdahulu pernah mendapatkan beban yang sangat berat, dan hamba memohon agar mereka dijauhkan dari beban yang serupa. Hamba juga memohon agar Allah tidak memikulkan kepada mereka apa yang tidak sanggup mereka pikul. Ini adalah pengakuan akan kelemahan manusia dan permohonan agar Allah memberikan beban yang sesuai dengan kemampuan

mereka.”

Ayat ini diakhiri dengan permohonan maaf, ampunan, dan rahmat dari Allah, serta permohonan pertolongan dalam menghadapi kaum yang kafir. Hamba memohon agar Allah senantiasa melimpahkan rahmat dan perlindungan-Nya, serta memberikan kekuatan dalam menghadapi tantangan dari kaum yang tidak beriman. Dari ayat di atas penulis menyimpulkan bahwa Allah SWT tidak membebani seseorang melainkan sesuai dengan kemampuannya. Setiap individu akan mendapatkan balasan sesuai dengan usaha dan perbuatannya, baik itu kebaikan maupun keburukan. Dalam konteks pilihan dan ikhtiar, ayat ini menunjukkan bahwa manusia diberikan kebebasan dan kemampuan untuk memilih serta bertindak, dan mereka bertanggung jawab atas pilihan-pilihan tersebut. Permohonan dalam ayat ini juga mencerminkan pengakuan atas keterbatasan manusia dan harapan akan ampunan serta rahmat Allah

Di lingkungan sekolah, keberadaan siswa yang memiliki minat dan motivasi tinggi menjadi dambaan setiap guru. Kehadiran siswa dengan semangat belajar yang kuat memudahkan para guru dalam membimbing jalannya proses pembelajaran. Selain itu, pencapaian hasil belajar yang maksimal lebih mungkin tercapai karena siswa memiliki inisiatif serta dorongan dari dalam diri untuk meraih keberhasilan. (Ricardo 2020).

Menurut Muhibbin Syah (2021), faktor psikologis seperti minat, kecerdasan, bakat, dan kemampuan kognitif berperan penting dalam memengaruhi hasil belajar. Pendapat ini didukung oleh Silvia dan Ricardo, yang menjelaskan bahwa siswa dengan tingkat minat belajar yang tinggi cenderung mencapai prestasi

akademik yang baik, memiliki kebiasaan belajar yang terorganisir, mampu memahami materi secara mendalam, memiliki efikasi diri yang tinggi, serta menunjukkan performa belajar yang optimal. Sebaliknya, siswa dengan minat belajar rendah sering kali menunjukkan perilaku seperti menarik diri, absen dari sekolah, putus sekolah, mengalami kecemasan berlebihan, dan mendapatkan hasil akademik yang kurang memuaskan. (Ricardo, 2020).

Berdasarkan berbagai pendapat, dapat disimpulkan bahwa minat adalah salah satu faktor utama yang memengaruhi hasil belajar siswa. Sofan Amri menyatakan bahwa minat tercermin melalui ungkapan atau tindakan yang menunjukkan preferensi siswa terhadap sesuatu dibandingkan dengan hal lain, serta terlihat dari partisipasi mereka dalam aktivitas tertentu. Siswa yang memiliki minat terhadap suatu bidang biasanya lebih terlibat dalam proses pembelajaran dan mencapai hasil yang lebih baik. Minat ini juga berperan penting saat siswa memilih jurusan atau program studi. Oleh karena itu, sebaiknya siswa diberi kebebasan untuk memilih jurusan sesuai dengan minat yang dimilikinya. (Sofan Amri, 2022). (Baharuddin et al., 2022).

Secara umum, siswa akan dihadapkan pada berbagai pilihan jurusan yang harus dipilih salah satunya. Proses pemilihan jurusan ini tidaklah mudah, karena keputusan tersebut akan berdampak pada hasil pendidikan dan perjalanan karier di masa depan. Oleh karena itu, penting bagi siswa untuk memahami minat mereka dan menentukan jurusan yang sesuai dengan keinginan serta potensi yang dimiliki.

Sistem rekomendasi menjadi topik yang semakin penting dalam dunia TI dan bisnis karena dapat memberikan rekomendasi yang dipersonalisasi dan akurat

kepada pengguna. Di era data yang terus meningkat, sistem rekomendasi menjadi semakin efektif dan efisien dalam mengolah data dan menghasilkan rekomendasi yang relevan. Ada berbagai jenis sistem rekomendasi, seperti *Content Based Filtering*, *Collaborative Filtering*, dan *Hybrid Filtering* (kombinasi dari *Content Based Filtering* dan *Collaborative Filtering*) (Zhang et al., 2019).

Sistem rekomendasi berbasis *content-based filtering* adalah sistem yang memberikan rekomendasi dengan mempertimbangkan kesamaan atribut antara item atau produk yang diminati (Fajriansyah et al., 2021). Sistem ini bekerja dengan membandingkan atribut item dengan atribut yang terdapat pada profil pengguna untuk menghasilkan rekomendasi. Algoritma ini dirancang untuk memberikan rekomendasi berdasarkan informasi yang telah dikumpulkan mengenai pengguna. Pendekatan ini memanfaatkan kesamaan dalam produk, layanan, atau fitur kontennya, serta memproses data pengguna untuk menyajikan rekomendasi yang relevan (Huda et al., 2022).

Content-Based Filtering digunakan karena algoritma ini tidak memerlukan data dari pengguna lain untuk menghasilkan rekomendasi. Pendekatan ini sangat bermanfaat ketika ada pengguna baru yang bergabung dengan sistem atau ketika data pengguna lain tidak tersedia. Selain itu, algoritma ini memungkinkan sistem memberikan rekomendasi yang akurat dan sesuai dengan preferensi pengguna.

Pada penelitian terdahulu yang dilakukan oleh Larasati & Februariyanti (2021) pada sistem rekomendasi produk Emina Cosmetics dengan menggunakan teknik penyaringan berbasis konten (*content-based filtering*) biasanya digunakan dalam sistem rekomendasi untuk memberikan rekomendasi. Artikel, berita, atau

website. Metode ini mengevaluasi data tentang item dan membandingkannya dengan data tentang item yang diakses atau dipilih pengguna. Dengan menggunakan teknik Collaborative Filtering berbasis konten, penelitian ini merekomendasikan 10 produk dengan skor kemiripan tertinggi sebesar 0,7195 berdasarkan perhitungan cosine similarity. Pada penelitian yang dilakukan oleh Rifaldy & Setiawan (2022) mengenai Collaborative Filtering pada sistem rekomendasi film Netflix dengan menggunakan Collaborative Filtering menggunakan algoritma Weighted Slope One Twitter, Collaborative Filtering menghasilkan nilai MAE yang relatif rendah yakni sebesar 0,924 pada $N=5$. Jika kita melakukan percobaan pada dataset yang terdiri dari 6183 dataset dengan 35 pengguna dan 791 film. Collaborative filtering juga terbukti mampu meminimalisir jumlah data kosong, namun tidak dapat menyelesaikan semua data kosong. Weighted Slope One juga terbukti sangat baik dalam memprediksi nilai evaluasi, dan dapat menghasilkan nilai MAE yang relatif rendah yaitu 0,568 pada $N=10$. Dapat disimpulkan bahwa sistem bekerja dengan baik meskipun nilai MAE yang dihasilkan rendah pada dataset yang diperoleh dari Twitter.

Ada dua jenis CF, CF berbasis memori dan CF berbasis model. CF berbasis memori menggunakan interaksi pengguna dengan item untuk menghitung kesamaan antara pengguna atau item, sehingga memungkinkannya merekomendasikan item yang sesuai dengan preferensi yang diperoleh. CF berbasis model, sebaliknya, menggunakan model matematika atau algoritma pembelajaran mesin untuk membuat rekomendasi (Prayoga & Kusnawi, 2022). CF berbasis model memiliki beberapa keunggulan dibandingkan CF berbasis memori

Meningkatkan skalabilitas dan mengatasi masalah ketersebaran data dengan lebih baik. Dalam CF berbasis model, algoritma dekomposisi nilai tunggal (SVD) banyak digunakan untuk menganalisis interaksi pengguna dengan item dan membuat rekomendasi (W. Zhang et al., 2019).

Salah satu masalah yang sering dialami oleh siswa kelas 12 saat menghadapi kelulusan adalah kebingungan dalam menentukan program studi dan universitas yang sesuai dengan minat serta kemampuan mereka. Banyak siswa yang belum memiliki gambaran jelas mengenai konsentrasi atau minat studi untuk melanjutkan pendidikan ke jenjang berikutnya. Bahkan, siswa yang sudah memiliki minat tertentu pun masih sering merasa bingung dalam memilih program studi dan perguruan tinggi yang akan mereka tempuh..

Oleh karena itu, penelitian ini bertujuan menggunakan *content Based filtering* untuk memberikan rekomendasi pemilihan program studi dan perguruan tinggi kepada siswa. Metode ini di pilih untuk mengamati dalam memberikan rekomendasi kepada siswa.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas , rumusan masalah yang dapat di ambil dalam penelitian ini adalah:

1. Bagaimana cara merancang sistem rekomendasi menggunakan metode content-based filtering untuk memberikan saran dalam pemilihan program studi dan perguruan tinggi bagi siswa
2. Bagaimana cara menilai tingkat keakuratan dari sistem rekomendasi yang telah dikembangkan

1.3 Batasan Penelitian

Penelitian ini memiliki Batasan sebagai berikut:

1. Sistem rekomendasi untuk pemilihan program studi hanya berfungsi sebagai panduan bagi siswa, sedangkan keputusan akhir terkait pemilihan program studi tetap sepenuhnya merupakan kewenangan siswa..
2. Metode yang digunakan dalam penelitian ini adalah metode *content-based filtering* dengan memberikan rekomendasi pada minat siswa

1.4 Tujuan Penelitian

Tujuan utama dari penelitian ini adalah memberikan rekomendasi padaminat siswa untuk memilih program studi dengan menggunakan penerapan metode *content-based filtering*.

1.5 Manfaat Penelitian

Penulis mengharapkan hasil penelitian ini dapat memberikan manfaat sebagai berikut:

1. Membantu mahasiswa dalam proses pemilihan program studi
2. Memberikan sumbangan terhadap kemajuan ilmu pengetahuan dan teknologi, serta memperluas penelitian untuk pengembangan di masa depan.

BAB II

STUDI PUSTAKA

2.1 Sistem Rekomendasi

Sistem rekomendasi adalah sebuah sistem yang dirancang untuk memberikan saran kepada pengguna sebagai alternatif pilihan dalam pengambilan keputusan sesuai preferensi mereka. Salah satu algoritma yang sering digunakan dalam sistem ini adalah content-based filtering (Miquido, 2020). Algoritma ini bekerja dengan memberikan rekomendasi berdasarkan kemiripan atribut antara item atau barang dengan atribut profil pengguna. Keunggulan dari content-based filtering meliputi ketidakbergantungan pada data pengguna lain, kemampuan untuk merekomendasikan item secara spesifik sesuai dengan preferensi individu, serta tidak memerlukan data historis dari pengguna (Nastiti, 2019).

Sistem rekomendasi adalah suatu sistem atau aplikasi yang dibuat untuk menyediakan dan memberikan rekomendasi mengenai item-item bagi pengguna sistem untuk mengambil keputusan yang diinginkan (Februariyanti et al., 2021). Sistem rekomendasi membantu pengguna menemukan dan memilih item (buku, film, produk, toko, restoran, dll.) dari sejumlah besar data yang tersedia di Internet dan sumber informasi elektronik lainnya. Mengingat kumpulan data yang besar dan deskripsi kebutuhan pengguna, sistem merekomendasikan sejumlah item kepada pengguna yang cocok dengan deskripsi tersebut.

Sistem rekomendasi adalah sebuah perangkat lunak dan pendekatan yang memberikan saran terkait item yang sesuai untuk pengguna. (Ricci et,al 2020).Saran yang diberikan Terkait dengan berbagai pilihan yang dibuat oleh pengguna, seperti

barang yang ingin dibeli, musik yang akan diputar, atau bacaan yang sesuai untuk dibaca.

Sistem rekomendasi akan memberikan berbagai saran untuk setiap pengguna, bukan sekadar menampilkan daftar item yang relevan. Oleh karena itu, setiap pengguna akan mendapatkan rekomendasi yang unik, yang disesuaikan dengan profil dan preferensi mereka.

Namun, untuk memberikan rekomendasi yang bersifat personal, sistem perlu memiliki pemahaman tentang pengguna. Setiap sistem rekomendasi harus mampu mengembangkan dan memelihara model atau profil pengguna yang mencakup, misalnya, informasi terkait minat atau preferensi pengguna.. (Rajaraman et,al 2020).

Menurut Putri dkk (2022) dalam Tinjauan Pustaka Sistem Rekomendasi Buku dan Film Sistem rekomendasi dapat memberikan solusi dalam situasi dimana pengguna kebingungan dalam memilih artikel yang sesuai dengan selera. Dengan menggunakan data pengguna yang telah diolah, sistem rekomendasi dapat memberikan rekomendasi spesifik dan personal yang lebih disesuaikan dengan selera dan preferensi pengguna. Oleh karena itu, sistem rekomendasi sangat membantu pengguna untuk memilih item yang tepat dan sesuai dengan keinginannya. Ini menghemat waktu dan tenaga pengguna dalam mencari judul yang sesuai dengan selera mereka. Tujuan dari sistem rekomendasi adalah untuk memperkirakan preferensi pengguna terhadap item dan secara aktif merekomendasikan item yang kemungkinan besar disukai pengguna. Model rekomendasi umumnya diklasifikasikan ke dalam tiga kategori: pemfilteran

kolaboratif, pemfilteran berbasis konten, dan pemfilteran hibrid (Zhang et al., 2019).

2.2 Content-Based Filtering

Content-based filtering merupakan sebuah algoritma yang digunakan dalam sistem rekomendasi untuk menyarankan item yang memiliki kemiripan dengan item yang sebelumnya disukai pengguna, berdasarkan fitur atau atribut tertentu. Pada pendekatan ini, setiap item dilengkapi dengan label atau atribut yang menjelaskan karakteristiknya. Sebagai contoh, dalam konteks film, atribut tersebut dapat berupa genre, aktor utama, sutradara, atau kata kunci dari deskripsi film. Selanjutnya, setiap pengguna memiliki profil yang menggambarkan preferensi mereka terhadap atribut-atribut tersebut. Profil ini dapat dibuat berdasarkan riwayat interaksi pengguna dengan item sebelumnya atau berdasarkan data yang disediakan pengguna, seperti pilihan atau penilaian mereka (Roy, 2020).

Sistem rekomendasi yang menerapkan metode content-based melakukan pembelajaran untuk memberikan saran item yang serupa dengan item yang sebelumnya disukai atau dipilih oleh pengguna. Perhitungan kemiripan antar item dilakukan berdasarkan fitur-fitur yang ada pada item yang dibandingkan (Ricci et al., 2020). Dengan demikian, metode ini tidak bergantung pada apakah item tersebut baru (belum pernah dipilih oleh pengguna lain) atau sudah dikenal. Sebagai contoh, dalam sistem rekomendasi untuk pemilihan program studi siswa, jika siswa kesulitan memilih program studi, sistem akan mencoba memberikan rekomendasi program studi yang kemungkinan besar akan disukai oleh siswa tersebut.

Metode ini biasanya digunakan untuk penyaringan informasi dalam sistem yang berbasis teks, tetapi dalam beberapa tahun terakhir, metode ini semakin populer diterapkan pada berbagai kasus lain, untuk mengatasi keterbatasan yang ada pada sistem rekomendasi yang menggunakan metode lain. Salah satu keunggulannya adalah kemampuannya untuk mengatasi kelemahan dari collaborative filtering yang tidak dapat memberikan rekomendasi untuk item-item baru.

Metode ini, yang dikenal sebagai metode berbasis konten, berasal dari disiplin ilmu retrieval informasi (sistem pencarian informasi) dan filtering informasi (Adomavicius et al., 2018). Penelitian yang berkaitan dengan retrieval informasi dan filtering informasi pada awalnya lebih banyak berfokus pada teks sebagai objek kajian, sehingga metode berbasis konten sering digunakan dalam sistem yang berorientasi pada teks.

Konten yang dipelajari dalam metode ini umumnya berupa informasi fitur tentang item yang ada (berbasis item). Namun, seiring berjalannya waktu, profil pengguna juga dapat dijadikan sebagai konten yang dianalisis untuk mencari kesamaan (berbasis pengguna). Sebagai contoh, pada toko buku online, jika profil pengguna A mirip dengan profil pengguna B (misalnya, keduanya melihat, menyukai, dan membeli buku yang sama), maka sistem rekomendasi akan berusaha menyarankan buku yang disukai oleh pengguna A kepada pengguna B, dan sebaliknya.

2.3 Minat Siswa

a. Pengertian Minat Siswa

Menurut kamus besar Bahasa Indonesia (KBBI), adalah kecenderungan, nafsu, atau Hasrat yang kuat terhadap sesuatu.(Anna rufaida 2019). minat adalah proses pengembangan yang menggabungkan semua keterampilan yang ada untuk membimbing orang ke dalam aktivitas yang diminati.(Linda Mey Lestari 2020).

Minat adalah suatu dorongan atau faktor dalam pikiran seseorang yang secara efektif membangkitkan minat atau perhatian, sehingga mengarah pada pemilihan objek atau kegiatan yang bermanfaat, menyenangkan, dan lama kelamaan mendatangkan kepuasan bagi orang tersebut.(Ahmad Susanto 2019).

(Menurut Sudirman 2020) adalah suatu keadaan yang terjadi Ketika seseorang mengenali ciri-ciri dan makna temporal suatu keadaan dalam hubungan dengan keinginan dan kebutuhannya sendiri.

Menurut kamus Psikologi karya J.P. Chaplin, dijelaskan bahwa minat (interest) adalah (AnnaRufaida 2019).

- 1) Sikap yang berlangsung secara berkesinambungan membentuk pola perhatian seseorang, sehingga membuatnya menjadi lebih selektif terhadap objek yang menarik minatnya.
- 2) Perasaan yang menunjukkan bahwa suatu aktivitas, pekerjaan, atau objek memiliki nilai atau makna bagi seseorang.
- 3) Satu kondisi motivasi atau kumpulan dorongan yang mengarahkan.

Berdasarkan pengertian tersebut, dapat disimpulkan bahwa minat adalah kecenderungan atau ketertarikan seseorang terhadap sesuatu yang, ketika terpenuhi, memberikan kepuasan tersendiri tanpa adanya paksaan. Seseorang akan melakukan hal yang diminati dengan senang hati, dan aktivitas tersebut sering menjadi bagian dari kesehariannya. Jika dikaitkan dengan minat siswa dalam memilih jurusan, siswa akan memilih jurusan berdasarkan rasa ketertarikannya sendiri tanpa tekanan dari pihak lain, sehingga mereka akan merasa lebih antusias dan menikmati pembelajaran mata pelajaran yang ada di jurusan tersebut..

b. Jenis – jenis Minat

Minat dapat dikelompokkan ke dalam beberapa jenis, yaitu:

- 1) Minat vokasional yang berfokus pada bidang pekerjaan tertentu.
 - a) Minat professional meliputi ketertarikan pada ilmu pengetahuan, seni, dan kesejahteraan sosial.
 - b) Minat komersial mencakup pekerjaan di dunia usaha seperti perdagangan, perikanan, akuntansi, kesekretariatan, dan sebagainya
 - c) Minat kegiatan fisik, seperti ketertarikan pada aktivitas mekanis, fisik, luar ruangan, dan lainnya.
- 2) Minat avokasional yaitu minat yang bertujuan untuk mendapatkan kepuasan pribadi atau sebagai hobi, seperti petualangan, hiburan, apresiasi seni, ketelitian, dan sebagainya.

Menurut Kuder dalam (ahmad Susanto 2019), minat dapat dikelompokkan menjadi 10 jenis:

- a) Minat terhadap alam sekitar, yaitu ketertarikan pada pekerjaan yang berhubungan dengan alam, hewan, dan tumbuhan.
- b) Minat mekanis yakni ketertarikan pada pekerjaan yang berkaitan dengan mesin atau alat mekanik.
- c) Minat hitung menghitung, yaitu ketertarikan pada pekerjaan yang membutuhkan perhitungan.
- d) Minat terhadap ilmu pengetahuan, yaitu keinginan untuk menemukan fakta baru dan memecahkan masalah.
- e) Minat persuasif, yakni ketertarikan pada pekerjaan yang melibatkan kemampuan memengaruhi orang lain
- f) Minat seni, mencakup ketertarikan pada pekerjaan yang berkaitan dengan seni, kerajinan, dan kreativitas.
- g) Minat literer, yaitu ketertarikan pada aktivitas membaca dan menulis berbagai jenis karya.
- h) Minat musik meliputi ketertarikan pada aktivitas terkait musik, seperti menonton konser atau memainkan alat musik.
- i) Minat layanan sosial, yaitu ketertarikan pada pekerjaan yang berfokus membantu orang lain.
- j) Minat klerikal, mencakup ketertarikan pada pekerjaan administratif

c. Ciri – ciri Minat

Menurut Elizabeth Hurlock, terdapat tujuh karakteristik minat yang tidak dibedakan berdasarkan sifatnya, apakah muncul secara spontan atau terbentuk secara terstruktur, sebagaimana dijelaskan oleh Gagne.

Karakteristik tersebut adalah sebagai berikut (Ahmad Susanto 2019).

- 1) Minat berkembang seiring dengan pertumbuhan fisik dan mental seseorang.
- 2) Minat seseorang sangat dipengaruhi oleh kegiatan belajar. Salah satu faktor yang meningkatkan minat adalah adanya kesiapan untuk belajar.
- 3) Minat juga bergantung pada peluang untuk belajar. Kesempatan belajar menjadi faktor yang sangat bernilai karena tidak semua orang dapat merasakannya.
- 4) Perkembangan minat bisa saja memiliki batasan tertentu.
- 5) Faktor budaya turut memengaruhi minat seseorang.
- 6) Minat memiliki muatan emosional
- 7) Minat juga bersifat egosentris, yang berarti ketika seseorang menyukai sesuatu, keinginan untuk memilikinya pun akan muncul

d. Fungsi Minat

Minat berhubungan erat dengan sikap kebutuhan seseorang dan mempunyai fungsi sebagai berikut

- 1) Sumber motivasi yang kuat untuk belajar. Anak yang berminat terhadap sebuah kegiatan baik permainan maupun pekerjaan akan berusaha lebih keras untuk belajar dibandingkan anak yang kurang berminat
- 2) Minat memengaruhi bentuk intensitas apresiasi anak. Ketika anak mulai berfikir tentang pekerjaan mereka di masa yang akan datang, semakin besar minat mereka terhadap kegiatan di kelas atau luar kelas yang mendukung tercapainya aspirasi itu

- 3) Menambah kegairahan pada setiap kegiatan yang ditekuni seseorang, anak yang berminat terhadap suatu pekerjaan atau kegiatan, pengalaman mereka jauh lebih menyenangkan daripada mereka yang merasa bosan.

e. Faktor yang mempengaruhi Minat Siswa

Terdapat dua faktor yang mempengaruhi minat siswa, yaitu (Ade Tri Rahmadani et.,al 2020)

- 1) Faktor internal merupakan faktor yang berasal dari dalam diri siswa mencakup:
 - a) Faktor fisik seperti kondisi kesehatan dan keberadaan cacat tubuh
 - b) Faktor psikologis meliputi perhatian, bakat, motivasi, tujuan, serta kesiapan siswa
 - c) Faktor kelelahan yang meliputi kelelahan fisik dan mental.
- 2) Faktor eksternal merupakan faktor yang berasal dari luar diri siswa, yang meliputi:
 - a) Faktor keluarga mencakup pola pengasuhan orang tua, hubungan antar anggota keluarga, suasana di rumah, dan latar belakang budaya keluarga
 - b) Faktor sekolah seperti metode pengajaran guru, interaksi antara guru dan siswa, serta fasilitas yang tersedia di sekolah.
 - c) Faktor Masyarakat termasuk hubungan siswa dengan masyarakat serta interaksi dengan teman sebaya.

2.4 Penelitian Terdahulu

Pada penelitian yang berjudul “ Rekomendasi Program Studi untuk Siswa SMA Sederajat Menggunakan Metode Hybrid Recommendation dengan Content-Based filtering dan Collaborative filtering” yang diteliti oleh (Muthi Ishlah et.al 2020) Tujuan dalam penelitian ini adalah untuk membangun sistem rekomendasi pemilihan program studi untuk siswa/i SMA sederajat menggunakan rekomendasi hybrid dengan mengembangkan content-based filtering dan collaborative filtering. Pendekatan yang digunakan dalam content-based filtering adalah *naïve bayes* dan collaborative filtering adalah item-based collaborative filtering. Pertama, data yang diperoleh berupa nilai fisika, matematika, bahasa Indonesia dan bahasa Inggris akan diolah dengan *naïve bayes* sehingga mendapatkan program studi dengan probabilitas tertinggi. Kemudian metode item based-collaborative filtering dibangun dan diterapkan pada sistem rekomendasi dengan menghitung rating program studi alternatif yang akan diberikan kepada siswa sesuai dengan rating yang diberikan sebelumnya. Selain itu, penelitian ini juga bertujuan untuk mengetahui akurasi yang dihasilkan oleh sistem rekomendasi yang telah dibangun dengan rekomendasi *hybrid* dengan menghitung akurasi *naïve bayes* dan mean absolute error dari item based-collaborative filtering.

Dalam penelitian yang berjudul “Sistem rekomendasi mata kuliah pilihan mahasiswa content-based filtering dan collaborative filtering (studi kasus universitas Brawijaya)” yang diteliti oleh (Aditya Fitri Hananta. Putra 2019) Penelitian ini bertujuan untuk merancang dan mengembangkan sistem rekomendasi dengan menggunakan metode collaborative filtering, yang dapat memberikan saran

pemilihan mata kuliah berdasarkan mata kuliah yang telah diambil sebelumnya secara akurat. Selain itu, penelitian ini juga bertujuan untuk merancang dan mengembangkan sistem rekomendasi yang berbasis content-based filtering, yang mampu memberikan rekomendasi dalam pemilihan mata kuliah. baru yang belum diambil sebelumnya secara akurat, serta mengevaluasi tingkat keakuratan atau kelayakan dari sistem rekomendasi yang dibangun. Dalam penelitian yang berjudul ” Sistem rekomendasi objek wisata di pontianak berbasis android menggunakan metode content- based filtering” yang diteliti oleh (Kevin christofer 2020) tujuan dalam penelitian ini adalah membantu dengan wisatawan dengan membangun sistem rekomendasi pariwisata yang berbasis android serta mempermudah para wisatawan melakukan pencarian tempatwisata menggunakan sistem rekomendasi pariwisata .

Dalam penelitian yang berjudul “Sistem rekomendasi kedai kopi dengan menggunakan metode collaborative filtering di kota yogyakarta berbasis web “ yang di teliti oleh (Bambang Hermanto 2020) tujuan dalam penelitian ini adalah mengembangkan sistem rekomendasi kedai kopi dengan metode *collaborative filtering* di kota yogyakarta dengan tujuan merekomendasikan kedai kopi favorit oleh para pecinta kopi berdasarkan rating angka .

Dalam penelitian yang berjudul “Sistem rekomendasi content-based filtering menggunakan TF-IDF vector similarity untuk rekomendasi artikel berita” yang di teliti oleh (Arif akbarul huda 2022) tujuan dalam penelitian ini adalah untuk membantu menemukan informasi yang relevan di kembangkanlah sistem rekomendasi . sistem rekomendasi akan membantu mengistemasi nilai prediksi atau

relevansi dari sebuah artikel dan kemudian membuat rangking atas artikel- artikel tersebut yang sesuai dengan minat pengguna.

Dalam penelitian yang berjudul “Perbandingan algoritma content-based filtering dan collaborative filtering dalam rekomendasi kegiatan ekstrakurikuler siswa” yang di teliti oleh (Dio Sukma Pradana et.al 2022) tujuan dalam penelitian ini adalah untuk menguji dan membandingkan kinerja model content-based filtering dan collaborative filtering berbasis naïve bayes dalam merekomendasikan kegiatan ekstrakurikuler yang tepat bagi siswa. Pengujian masing-masing model dilakukan dengan memabgi data latih dan data uji dalam perbandingan 80% dan 20% . proses pelatihan menggunakan library recommender NET.

Tabel 2.1 Penelitian Terdahulu

Referensi	Judul	Metode Penelitian	Hasil Penelitian	Perbedaan
(Muthi Ishlah et.al, 2020)	Sistem Rekomendasi Program Studi untuk Siswa SMA Sederajat Menggunakan Metode Hybrid Recommendation dengan Content Based Filtering dan Collaborative Filtering	Metode yang digunakan untuk memberikan rekomendasi program studi adalah hybrid recommendation.	Hasil penelitian ini menunjukkan bahwa pengolahan data nilai dengan naïve bayes menghasilkan akurasi sebesar 88.7% ,	Perbedaannya adalah jurnal ini menggunakan metode hybrid recommendation
(Aditya Fitri Hananta. Putra, 2019)	Sistem rekomendasi mata kuliah pilihan mahasiswa content-based filtering dan collaborative filtering (studi kasus Universitas Brawijaya)	Metode yang di gunakan content-based filtering dan collaborative filtering	Berdasarkan hasil pengujian didapatkan hasil akurasi berkisar antara 10,98% hingga 34,69% untuk pengujian terhadap semester yang tidak saling berkesesuaian dengan metode content-based filtering 6,3% hingga 42,12%	Perbedaannya adalah penelitian ini menggunakan data mahasiswa untuk memilih matakuliah sedangkan penelitian saya menggunakan data siswa untuk memilih program studi

Referensi	Judul	Metode Penelitian	Hasil Penelitian	Perbedaan
(Kevin Christofer, 2020)	Sistem rekomendasi objek wisata di pontianak berbasis andorid menggunakan metode content-based filtering	Metode yang digunakan adalah content-based filtering	Hasil dari pembuatan sistem ini juga dapat menjadi meningkatkan minat wisatawan berwisata ke kota pontianak	Jadi perbedaanya adalah penelitian ini menggunakan objeknya wisata sedangkan penelitian saya menggunakan objek sekolah
(Bambang Hermanto, 2020)	Sistem rekomendasi kedai kopi dengan menggunakan metode collaborative filtering di kota yogyakarta berbasis web	Metode yang digunakan adalah collaborative filtering	dapat memberikan informasi mengenai kedai kopi yang sesuai dengan parameter yang telah disediakan oleh sistem rekomendasi kedai kopi dan dapat membantu parakonsumen atau para pecinta kopi untuk mencari kedai kopi berdasarkan parameter yang telah disediakan oleh sistem rekomendasi	Jadi perbedaanya penelitian ini menggunakan metode collaborative filtering sedangkan penelitian saya menggunakan content-based filtering
(Arif Akbarul Huda, 2022)	Sistem rekomendasi content-based filtering menggunakan TF-IDF vector similarity untuk rekomendasi artikel berita	Metode yang digunakan adalah content-based filtering	Hasil dari peneltian ini content-based filtering mampu mmeberikan score Recall@5 sekitar 73% dan Recall @10 sekitar 80%	Perbedaanya pada penelitian ini menggunakan populasi mahasiswa aktif prodi informatika universitas unikom yogyakarta pada semester ganjil 2021 sebanyak 3870 orang

Referensi	Judul	Metode Penelitian	Hasil Penelitian	Perbedaan
(Dio sukma pradana et.al 2022)	Perbandingan algoritma content-based filtering dan collaborative filtering dalam rekomendasi kegiatan ekstrakurikuler siswa	Metode yang digunakan adalah content-based filtering dan collaborative filtering	Hasil uji mode naïve bayes untuk content-based filtering menunjukkan tingkat akurasi 74% sedangkan collaborative filtering memperoleh 56%	Jadi perbedaannya adalah penelitian ini menggunakan model naïve bayes jenis gaussian

Jadi penelitian ini menggunakan metode content based filtering untuk sistem rekomendasi program studi, penelitian ini fokus pada program studi berdasarkan minat siswa ,menggunakan data seperti preferensi siswa,nilai akademik, dan data deskriptif dari program studi, sedangkan penelitian sebelumnya menggunakan metode collaborative filtering,*hybrid methods*,atau *rule-based system* .

Tujuan dari penelitian adalah untuk mengembangkan system rekomendasi yang membantu mahasiswa memilih program studi perguruan tinggi berdasarkan minatnya, system rekomendasi ini memberikan rekomendasi yang relevan dengan menggunakan teknik penyaringan berbasis konten yang fokus pada analisis konten program studi dan preferensi mahasiswa. Metode penyaringan berbasis konten fokus pada karakteristik dari item yang direkomendasikan dan membandingkan dengan preferensi atau minat pengguna(siswa). System menganalisis deskripsi dan atribut program seperti kursus yang ditawarkan, bidang studi, dan peluang karir dan mencocokkannya dengna minat dan preferensi siswa yang ditentukan melaluisurvei dan analisis data lainnya.

Perbedaan dari penelitian sebelumnya penelitian ini berbeda dari penelitian

sebelumnya dalam beberapa hal:

1. Pendekatan metode

Banyak system pemberi rekomendasi sebelumnya menggunakan pendekatan penyaringan kolaboratif berdasarkan data dari preferensi, dalam studi ini fokus pada pemfilteran berbasis konten, yang berarti bahwa system ini tidak bergantung pada data dari pengguna lain, tetapi hanya pada atribut dan karakteristik tertentu dari kursus, dan minat pribadi siswa.

2. Fokus pada minat siswa

Penelitian ini berfokus pada penyesuaian rekomendasi dengan minat khusus siswa sebelumnya banyak penelitian mungkin lebih umum untuk merancang program studi berdasarkan kriteria standar, seperti mata pelajaran utama atau umum, tanpa mempertimbangkan preferensi pribadi siswa secara mendalam.

3. Peningkatan detail fitur

Kajian ini memerlukan analisis lebih rinci mengenai karakteristik program pembelajaran seperti peluang karir spesifik, kursus yang sangat relevan, atau aspek unik lainnya yang belum tentu dipertimbangkan dalam penelitian sebelumnya

4. Penyesuaian individu

System ini memberikan rekomendasi yang lebih personal dan relevan dengan mengutamakan kepentingan pribadi siswa, tidak seperti system yang lebih umum yang tidak secara khusus mempertimbangkan kedalaman minat siswa.

5. Penggunaan metode content-based filtering

Penelitian ini berfokus pada metode content-based filtering untuk memberikan rekomendasi program studi kepada siswa SMA, sebaliknya penelitian oleh Muthi

Ishlah et al.(2020) Menggunakan metode *hybrid* yang menggabungkan content-based filtering dan collaborative filtering untuk tujuan yang sama.

6. Objek dan subjek penelitian

- a. Penelitian ini khusus mengkaji rekomendasi program studi bagi siswa SMA sederajat
- b. enggan fokus pada data nilai mata pelajaran sebagai input utama
- c. Aditya Fitri Hananta Putra (2019) melakukan penelitian tentang sistem rekomendasi untuk mata kuliah pilihan, dengan menggunakan data mahasiswa sebagai subjek penelitiannya.
- d. Kevin Christofer(2020) meneliti system rekomendasi objek wisata berbasis Andorid diPontianak, dengan objek wisata sebagai fokusnya.
- e. Bambang Hermanto(2020) mengembangkan system rekomendasi kedai kopi di Yiogyakartadengan menggunakan metode collaborativefiltering.

7. Fokus pada Preferensi siswa

- a. Penelitian ini menekankan pentingnya memahami minat dan bakatsiswa untuk memberikankomendasi yang akurat dan bermafaat.
- b. Dalam konteks penelitian ini juga menyoroti pentingnya penyesuaian rekomendasi denganpreferensi pengguna, namundengna fokus pada artikel berita.

Kesimpulan dari penelitian ini bertujuan untuk menciptakan sistem rekomendasi yang lebih personal dan menyasar minat individu siswa dengan menggunakan Teknik content-based filtering. Ini merupakan kemajuan dibandingkan pendekatan yang lebih umum atau pendekatan yang mengandalkan

data pengguna lain dengan memberikan rekomendasi yang lebih akurat dan relevan berdasarkan karakteristik kursus tertentu dan preferensi pribadi siswa, penelitian ini akan memberikan kontribusi baru dalam bidang system rekomendasi dengan fokus pada pemilihan program studi bagi siswa SMA sederajat menggunakan metode content based filtering, perbedaan utama dengan penelitian sebelumnya terletak pada metode yang digunakan, objek dan subjek penelitian, serta fokus pada preferensi siswa dalam konteks Pendidikan. Dengan demikian diharapkan, penelitian ini dapat memberikan solusi yang efektif dan akurat untuk membantu siswa dalam memilih program studi yang sesuai dengan minat dan bakat mereka.

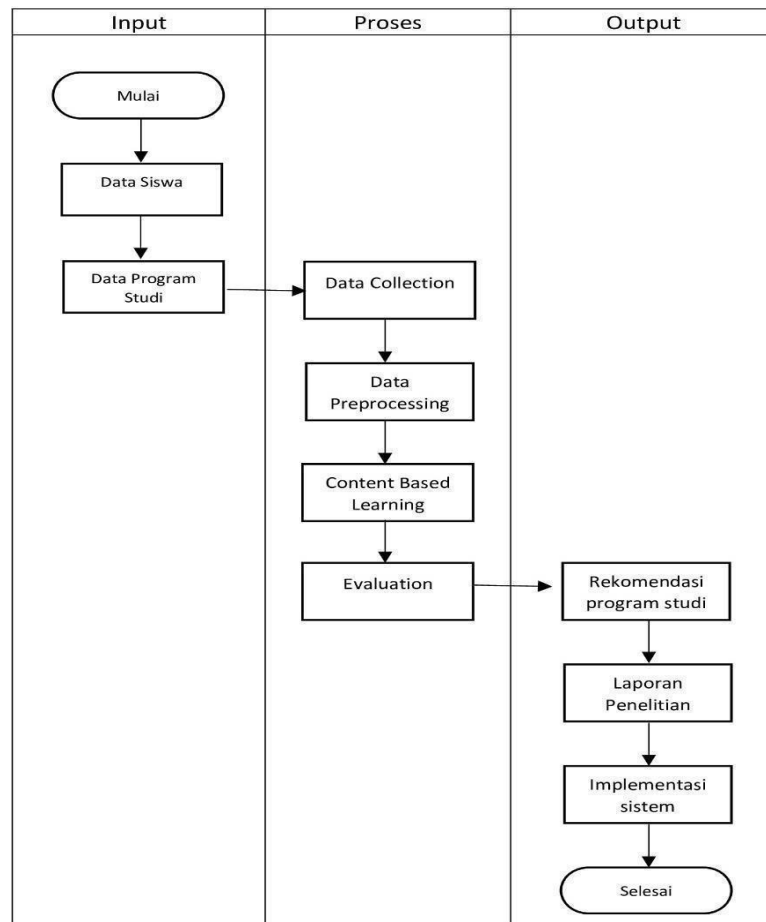
BAB III

METODE PENELITIAN DAN PERANCANGAN

Pada bab ini dijelaskan langkah-langkah yang akan diambil dalam menyusun penelitian. Tahapan penelitian ini dirancang untuk mengembangkan sistem rekomendasi yang mampu memberikan rekomendasi terkait minat siswa yang melakukan registrasi akademik menggunakan sistem rekomendasi.

3.1 Desain penelitian

Dibawah ini adalah prosedur penelitian yang akan dilakukan:



Gambar 3.1 Desain Penelitian

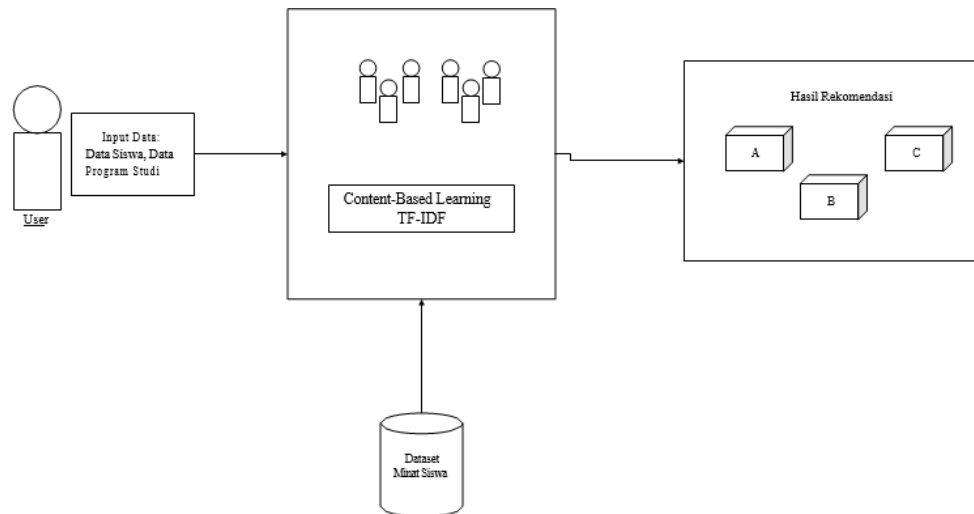
Desain penelitian sistem rekomendasi pemilihan program studi dengan menggunakan metode Content-Based Filtering dapat dirancang sebagai berikut. Tahap pertama adalah input, yang mencakup data siswa dan data program studi serta perguruan tinggi. Data siswa meliputi profil siswa (nama, usia, gender, sekolah, dll.), data akademik (nilai mata pelajaran, prestasi akademik, dll.), minat dan preferensi (jurusan yang diminati, bidang studi favorit, kegiatan ekstrakurikuler, dll.), serta hasil psikometrik (tes minat dan bakat, hasil tes kepribadian, dll.). Data program studi dan perguruan tinggi meliputi daftar program studi (nama program studi, fakultas, universitas, deskripsi, mata kuliah utama, dll.), data universitas (nama universitas, akreditasi, lokasi, fasilitas, biaya, dll.), dan data alumni (profil alumni, prestasi, pekerjaan saat ini, dll.). Selanjutnya adalah tahap proses, yang dimulai dengan pengumpulan data siswa dan data program studi dari berbagai sumber seperti survei dan database sekolah/universitas. Kemudian dilakukan preprocessing data yang mencakup cleaning (menghapus data yang tidak lengkap atau tidak relevan), normalization (menormalkan data agar seragam dan konsisten), dan feature extraction (mengidentifikasi fitur-fitur penting dari data yang akan digunakan dalam algoritma rekomendasi). Dalam metode Content-Based Filtering, profil siswa dibuat berdasarkan minat dan preferensi mereka, yang kemudian dibandingkan dengan atribut dari program studi yang ada. Kemiripan dihitung menggunakan metode seperti cosine similarity atau TF-IDF, menghasilkan daftar rekomendasi program studi yang sesuai dengan profil

siswa. Dalam metode Collaborative Filtering, data interaksi siswa dengan program studi dikumpulkan, misalnya program studi yang dipilih oleh siswa dengan profil serupa. Algoritma Collaborative Filtering (Memory-Based atau Model-Based) digunakan untuk menemukan pola dari data interaksi tersebut dan menghasilkan daftar rekomendasi berdasarkan pola yang ditemukan.

Tahap evaluasi dilakukan untuk mengukur keakuratan dan kelayakan rekomendasi menggunakan metrik evaluasi seperti Mean Absolute Error (MAE), Root Mean Square Error (RMSE), atau Precision dan Recall. Hasil rekomendasi dari kedua metode (Content-Based dan Collaborative Filtering) kemudian dibandingkan. Output dari penelitian ini mencakup rekomendasi program studi berupa daftar program studi yang direkomendasikan untuk masing-masing siswa berdasarkan hasil dari metode Content-Based Filtering, serta analisis perbandingan keakuratan dan kelayakan rekomendasi dari kedua metode. Laporan penelitian juga dihasilkan, yang menjelaskan proses penelitian, metode yang digunakan, dan hasil yang diperoleh, serta kesimpulan dan rekomendasi untuk implementasi sistem rekomendasi di sekolah atau lembaga pendidikan. Selain itu, prototipe atau aplikasi sistem rekomendasi dikembangkan untuk membantu siswa dalam memilih program studi dan perguruan tinggi, disertai panduan penggunaan sistem bagi siswa dan guru.

3.2 Desain Sistem

Gambaran sistem rekomendasi yang digunakan dalam penelitian ini adalah sebagai berikut:



Gambar 3.2 Desain Sistem

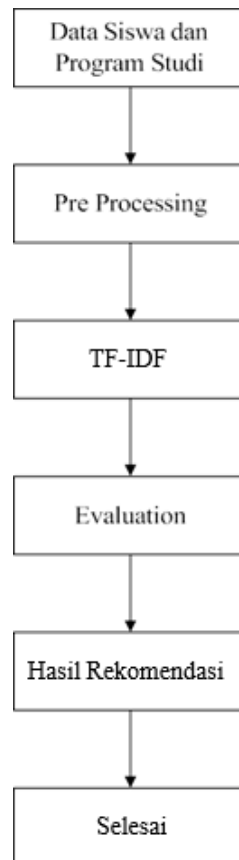
Desain sistem rekomendasi pemilihan program studi ini dimulai dengan tahap awal di mana siswa memulai interaksi dengan sistem. Tahap pertama adalah pendaftaran atau login, di mana siswa mendaftar atau login ke dalam sistem. Setelah itu, siswa mengisi profil dan data terkait dalam tahap input data siswa. Selanjutnya, data siswa diproses menggunakan metode Content-Based untuk menghasilkan rekomendasi yang sesuai. Sistem kemudian mengevaluasi hasil rekomendasi untuk memastikan akurasi. Setelah evaluasi, sistem menampilkan hasil rekomendasi kepada siswa. Proses selesai setelah siswa menerima rekomendasi. Desain sistem ini dirancang untuk memastikan bahwa proses dari input data hingga pemberian rekomendasi berjalan secara efektif dan efisien, memberikan siswa panduan yang dapat diandalkan dalam memilih program studi dan perguruan tinggi.

3.3 Analisis Pengolahan Data

Dalam analisis pengolahan data ini, langkah pertama adalah pengumpulan data dari berbagai sumber. Data siswa diperoleh melalui formulir pendaftaran dan survei yang mencakup informasi demografis, nilai akademik, minat, dan hasil tes psikometrik. Data program studi dan perguruan tinggi dikumpulkan dari database universitas, situs web resmi, dan publikasi akademik. Selanjutnya, dilakukan preprocessing data yang meliputi beberapa tahap. Tahap pertama adalah pembersihan data dengan menghapus data duplikat dan menangani nilai yang hilang melalui imputasi atau menghapus entri yang tidak lengkap. Langkah berikutnya adalah normalisasi data, yaitu menormalkan nilai akademik (misalnya, skala 0-100) dan menstandarkan format data untuk konsistensi (misalnya, format tanggal). Data kualitatif kemudian diubah menjadi bentuk numerik menggunakan teknik seperti one-hot encoding untuk kategori minat. Setelah preprocessing, dilakukan ekstraksi fitur dari data yang ada. Untuk data siswa, fitur yang diekstraksi meliputi demografi (usia, gender), nilai akademik (nilai rata-rata, nilai per mata pelajaran), minat (jurusan yang diminati, kegiatan ekstrakurikuler), dan hasil psikometrik (skor tes minat dan bakat, hasil tes kepribadian). Untuk data program studi, fitur yang diekstraksi meliputi atribut program studi (nama, deskripsi, mata kuliah utama, persyaratan masuk) dan data universitas (akreditasi, lokasi, fasilitas, biaya).

Dalam tahap Content-Based Filtering (CBF), profil siswa dibuat dalam bentuk vektor berdasarkan minat, nilai akademik, dan hasil psikometrik. Profil program studi juga dibuat dalam bentuk vektor atribut program studi. Kemiripan antara profil siswa dan profil program studi dihitung menggunakan metode seperti

cosine similarity atau TF-IDF. Berdasarkan kemiripan ini, dihasilkan daftar program studi yang memiliki skor kemiripan tertinggi dengan profil siswa. Dalam tahap Collaborative Filtering (CF), data interaksi siswa dengan program studi dikumpulkan, misalnya program studi yang dipilih oleh siswa lain dengan profil serupa. CF berbasis memori menghitung kemiripan antara pengguna atau antara item berdasarkan data interaksi, sedangkan CF berbasis model membangun model matematika atau menggunakan algoritma pembelajaran mesin seperti SVD untuk menganalisis interaksi dan membuat rekomendasi. Rekomendasi dihasilkan berdasarkan pola interaksi yang ditemukan. Tahap evaluasi menggunakan metrik evaluasi seperti Mean Absolute Error (MAE) yang mengukur rata-rata kesalahan absolut antara rekomendasi dan preferensi siswa sebenarnya, Root Mean Square Error (RMSE) yang mengukur rata-rata kesalahankuadrat antara rekomendasi dan preferensi siswa sebenarnya, serta Precision dan Recall yang mengukur akurasi dan kelengkapan rekomendasi yang diberikan. Hasil dari Content-Based Filtering kemudian dibandingkan untuk menentukan metode yang lebih efektif dalam memberikan rekomendasi kepada siswa.



Gambar 3.3 Analisis Pengolahan Data

3.4 Preprocessing Text

Preprocessing adalah tahapan awal dalam data mining untuk mengubah data mentah menjadi informasi yang lebih bersih dan bisa digunakan untuk pengolahan data selanjutnya. Pada tahapan preprocessing kata-kata atau teks yang tidak perlu atau tidak mempunyai arti akan dihilangkan, Dengan berkurangnya jumlah teks, sebagaimana itu meringankan proses selanjutnya dalam rangka menambang informasi yang berguna tanpa mengurangi arti ataupun makna serta informasi yang dikandung dalam dokumen tersebut. Pada penelitian ini dilakukan tahap preprocessing dengan memanfaatkan Natural Language Processing (NLP) dengan tools Natural Language Tool Kit (NLTK) yang digunakan untuk bahasa

pemrograman Python. Di antara Langkah - langkah yang terlibat dalam pra-pemrosesan adalah pelipatan kasus, tokenisasi, pemfilteran, dan stemming.

a. Case folding

Case folding adalah konversi karakter dari huruf besar ke huruf kecil (Achsani et al. 2023). Fungsi dari case folding adalah menyamaratakan penggunaan huruf di setiap dokumen agar lebih terstruktur dan konsisten. Dengan menghilangkan perbedaan antara huruf besar dan huruf kecil, analisis teks akan menjadi lebih mudah karena variasi tidak perlu lagi di pertimbangkan. contoh dari proses case folding dapat di lihat pada table

Tabel 3. 1 Case folding

Sebelum	Sesudah
Program Studi yang ditawarkan oleh Universitas ini sangat beragam. Beberapa di antaranya adalah Teknik Informatika, Psikologi, dan Ekonomi.	program studi yang ditawarkan oleh universitas ini sangat beragam. beberapadi antaranya adalah teknik informatika, psikologi, dan ekonomi.

Dalam contoh di atas, semua huruf dalam teks asli telah diubah menjadi huruf kecil, sehingga menghasilkan teks yang seragam untuk analisis lebih lanjut. Case folding membantu dalam mengurangi variasi dalam teks yang disebabkan oleh penggunaan huruf besar dan kecil, sehingga memungkinkan algoritma pemrosesan teks untuk bekerja lebih efektif.

b. Tokenizing

Tokenisasi, juga dikenal sebagai Analisis Leksikal, adalah proses mereduksi teks menjadi potongan-potongan yang lebih kecil dan lebih mudah dikelola yang dikenal sebagai token. Selain itu, pada tahap tokenizing juga dilakukan proses

penghilangan angka, teks, dan karakter lain yang tampaknya tidak berpengaruh pada penulisan ulang teks. Contoh dari proses *tokenizing* dapat dilihat pada table

Tabel 3. 2 Tokenizing

Sebelum	Sesudah
['Program', 'Studi', 'yang', 'ditawarkan', 'oleh', 'Universitas', 'ini', 'sangat',	['Program', 'Studi', 'yang', 'ditawarkan', 'oleh', 'Universitas', 'ini', 'sangat',

Dalam contoh di atas, teks asli telah dipecah menjadi kata-kata (token) individual, termasuk tanda baca seperti titik (.), dan koma (,). Tokenizing membantu dalam mempersiapkan teks untuk analisis lebih lanjut seperti stemming, lemmatization, dan analisis frekuensi kata.

c. Stopword removal

Stopword removal adalah penghapusan kata-kata umum yang rendah konten informasi (Kaur and Kaur Buttar 2018). Contoh dari *Stopword removal* dapat dilihat pada table

Tabel 3. 3 Stopword Removal

Sebelum	Sesudah
['Program', 'Studi', 'ditawarkan', 'Universitas', 'beragam', '.', 'Beberapa', 'antaranya', 'Teknik', 'Informatika', ',', 'Psikologi', ',', 'Ekonomi', '.']	['Program', 'Studi', 'ditawarkan', 'Universitas', 'beragam', '.', 'Beberapa', 'antaranya', 'Teknik', 'Informatika', ',', 'Psikologi', ',', 'Ekonomi', '.']

Dalam contoh di atas, stopwords seperti "yang," "oleh," "ini," "di," dan "adalah" telah dihapus dari teks asli. Proses stopwords removal ini membantu dalam menyaring kata-kata umum yang tidak signifikan, sehingga memungkinkan

analisis teks untuk lebih fokus pada kata-kata yang memiliki makna lebih penting.

d. Stemming

Stemming adalah proses untuk mengubah bentuk kata menjadi kata dasar. Imbuhan yang terdapat di awal kata (prefix) dan akhir kata (sufix) akan dihapus. Langkah ini bertujuan untuk mengkonsolidasikan berbagai bentuk kata yang berbeda ke dalam bentuk dasar, sehingga variasi kata dalam analisis berkurang. Contoh dari proses stemming dapat dilihat pada tabel.

Tabel 3. 4 Proses Stemming

Sebelum	Sesudah
['mahasiswa', 'tersebut', 'sedang', 'belajar', 'di', 'perpustakaan', '.', 'mereka', 'belajar', 'bagai', 'mata', 'ajar', 'yang', 'kait', 'dengan', 'teknik', 'informatika', '.']	['Mahasiswa', 'tersebut', 'sedang', 'belajar', 'di', 'perpustakaan', '.', 'Mereka', 'belajar', 'berbagai', 'mata', 'pelajaran', 'yang', 'berkaitan', 'dengan', 'teknik', 'informatika', '.']

Dalam contoh di atas, kata-kata seperti "mahasiswa-mahasiswa" diubah menjadi "mahasiswa", "berbagai" menjadi "bagai", dan "pelajaran" menjadi "ajar". Proses stemming ini membantu dalam mengkonsolidasikan berbagai bentuk kata yang berbeda menjadi satu bentuk dasar atau akar, sehingga memudahkan analisis teks lebih lanjut dengan mengurangi variasi kata.

3.5 Term Frequency (TF) dan Inverse Document Frequency (IDF)

TF-IDF adalah teknik yang digunakan dalam pemrosesan bahasa alam dan penambahan teks untuk mengevaluasi pentingnya suatu kata dalam sebuah dokumen relatif terhadap seluruh koleksi dokumen (korpus). TF-IDF adalah kombinasi dari dua metrik: Term Frequency (TF) dan Inverse Document Frequency (IDF).

1. **Term Frequency (TF):** Mengukur seberapa sering sebuah kata muncul dalam sebuah dokumen. TF dapat dihitung dengan rumus:

$$TF(t, d) = \frac{\text{Jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{Jumlah kata dalam dokumen } d}$$

2. **Inverse Document Frequency (IDF):** Mengukur seberapa penting sebuah kata dalam seluruh koleksi dokumen. IDF dapat dihitung dengan rumus:

$$IDF(t, D) = \frac{\text{Jumlah dokumen dalam koleksi } d}{\text{Jumlah dokumen yang mengandung kata } t}$$

3. **TF-IDF:** Merupakan hasil perkalian dari TF dan IDF:

$$TF - IDF(t, d, D) = TF(t, d) \times IDF(t, D)$$

Contoh perhitungan TF-IDF ditunjukkan pada Tabel 3.5 sebagai berikut.

Tabel 3.5 Perhitungan

Kata	TF Dokumen1	TF Dokumen2	TF Dokumen3	IDF	TF-IDF Dokumen1	TF-IDF Dokumen2	TF-IDF Dokumen3
Program	12-Jan	10-Jan	10-Jan	0	0	0	0
Studi	12-Jan	10-Jan	10-Jan	0	0	0	0
Teknik	12-Jan	10-Jan	0	0.176	0.0147	0.0176	0
Informatika	12-Jan	0	0	0.477	0.03975	0	0
Elektro	0	10-Jan	0	0.477	0	0.0477	0
Mencakup	0	10-Jan	0	0.477	0	0.0477	0
Tentang	0	10-Jan	0	0.477	0	0.0477	0
Elektronika	0	10-Jan	0	0.477	0	0.0477	0
Dan	12-Jan	10-Jan	0	0.176	0.0147	0.0176	0
Sistem	12-Jan	0	10-Jan	0	0	0	0
Informasi	12-Jan	0	10-Jan	0.176	0.0147	0	0.0176
Fokus	12-Jan	0	0	0.477	0.03975	0	0
Pada	12-Jan	0	0	0.477	0.03975	0	0
Pengembangan	12-Jan	0	0	0.477	0.03975	0	0
Perangkat	12-Jan	0	0	0.477	0.03975	0	0
Lunak	12-Jan	0	0	0.477	0.03975	0	0
Memadukan	0	0	10-Jan	0.477	0	0	0.0477
Teknologi	0	0	10-Jan	0.477	0	0	0.0477
Dengan	0	0	10-Jan	0.477	0	0	0.0477
Manajemen	0	0	10-Jan	0.477	0	0	0.0477
Bisnis	0	0	10-Jan	0.477	0	0	

3.6 Evaluasi

Dalam evaluasi sistem rekomendasi pemilihan program studi dengan menggunakan metode *Content-Based Filtering* dan *Collaborative Filtering*, beberapa metrik evaluasi yang dapat digunakan antara lain:

1. *Mean Absolute Error* (MAE):

MAE mengukur rata-rata kesalahan absolut antara rekomendasi yang diberikan oleh sistem dengan preferensi atau pilihan yang sebenarnya oleh siswa.

Semakin rendah nilai MAE, semakin baik kualitas rekomendasi yang diberikan oleh sistem. Misalkan setelah evaluasi, rata-rata MAE dari sistem Content-Based Filtering adalah 0.25 dan dari sistem Collaborative Filtering adalah 0.30. Semakin rendah nilai MAE, semakin baik kualitas rekomendasi sistem.

2. *Root Mean Square Error (RMSE):*

RMSE mengukur rata-rata kesalahan kuadrat antara rekomendasi sistem dengan preferensi siswa. RMSE memberikan gambaran tentang seberapa besar kesalahan yang terjadi dalam rekomendasi. Nilai RMSE yang lebih rendah menunjukkan kualitas rekomendasi yang lebih baik. Setelah perhitungan, RMSE dari sistem Content-Based Filtering adalah 0.35 dan dari sistem Collaborative Filtering adalah 0.40. Kembali, semakin rendah nilai RMSE, semakin baik kualitas rekomendasi sistem.

3. *Precision dan Recall:*

Precision mengukur seberapa akurat rekomendasi sistem, yaitu jumlah rekomendasi yang relevan dibandingkan dengan jumlah total rekomendasi yang diberikan. Recall mengukur seberapa lengkap sistem dalam memberikan rekomendasi yang relevan, yaitu jumlah rekomendasi yang relevan dibandingkan dengan jumlah keseluruhan item yang relevan. Semakin tinggi nilai precision dan recall, semakin baik kualitas rekomendasi sistem. Precision sistem Content-Based Filtering adalah 0.80, yang berarti 80% dari rekomendasi yang diberikan adalah relevan. Recall sistem Collaborative Filtering adalah 0.75, yang berarti 75% dari keseluruhan program studi yang relevan diprediksi dengan benar oleh sistem. Semakin tinggi nilai precision dan recall, semakin baik kualitas rekomendasi sistem.

4. Perbandingan Metode:

Selain itu, perbandingan antara metode Content- Based Filtering dan Collaborative Filtering dapat dilakukan berdasarkan hasil evaluasi menggunakan metrik-metrik di atas. Metode yang memberikan performa evaluasi yang lebih baik dapat dipilih untuk diterapkan dalam sistem rekomendasi. Berdasarkan evaluasi, sistem Content-Based Filtering memiliki performa yang sedikit lebih baik dalam hal MAE dan RMSE, sementara sistem Collaborative Filtering memiliki nilai precision yang sedikit lebih tinggi. Dengan demikian, kita dapat menyimpulkan bahwa kualitas rekomendasi dari kedua metode ini cukup serupa, tetapi bisa jadi lebih baik jika dikombinasikan dalam sebuah sistem hibrida.

3.7 Hasil Rekomendasi

Dalam hasil rekomendasi, sistem akan memberikan daftar program studi yang direkomendasikan untuk setiap siswa berdasarkan metode Content-Based Filtering. Ini adalah tahap akhir dari proses rekomendasi di mana siswa akan melihat rekomendasi yang diberikan oleh sistem. Contoh hasil rekomendasi dari metode ini bisa seperti berikut:

1. Siswa A:
 - a. Rekomendasi: Teknik Informatika
 - b. Alasan: Berdasarkan minat siswa dalam teknologi dan keterampilan pemrograman yang tinggi.
2. Siswa B:
 - a. Rekomendasi: Manajemen Bisnis
 - b. Alasan: Berdasarkan minat siswa dalam kepemimpinan dan

ke~~k~~eterampilan manajerial yang kuat.

BAB IV

HASIL DAN PEMBAHASAN

Bab ini membahas hasil implementasi serta pembahasan terkait sistem yang telah dirancang dalam bab metodologi penelitian. Beberapa subbab yang akan diuraikan mencakup pengumpulan data, implementasi basis data, tahap prapemrosesan, perhitungan *tf-idf*, perhitungan *cosine similarity*, hasil rekomendasi, implementasi sistem, serta pengujian.

4.1 Pengumpulan Data

Bab ini menyajikan hasil penelitian yang dilakukan berdasarkan dua sumberdata utama. Sumber data pertama adalah hasil pengumpulan minat dan bakat siswa SMAN 1 Sangkapura, yang diperoleh melalui kuesioner Google Form(Gform) yang diisi oleh siswa selama periode 19 September hingga 30 September 2024. Dalam pengumpulan data ini, sebanyak 100 siswa berpartisipasi dengan memberikan informasi terkait minat mereka dalam berbagai bidang studi. Kuesioner ini dirancang untuk mencakup beragam aspek, mulai dari minat akademik hingga hobi, dengan tujuan untuk menggali potensi siswa secara lebih mendalam.

Sumber data kedua berkaitan dengan program studi dari berbagai perguruan tinggi, yang diambil dari beberapa situs web resmi. Data ini mencakup 22 program studi yang mencerminkan berbagai disiplin ilmu, termasuk ilmu sosial, sains, teknologi, dan seni. Program studi ini dipilih sebagai referensi dalam analisis kesesuaian antara minat siswa dan program studi yang tersedia, sehingga diharapkan dapat memberikan gambaran yang lebih jelas mengenai pilihan yang

dapat diambil oleh siswa.

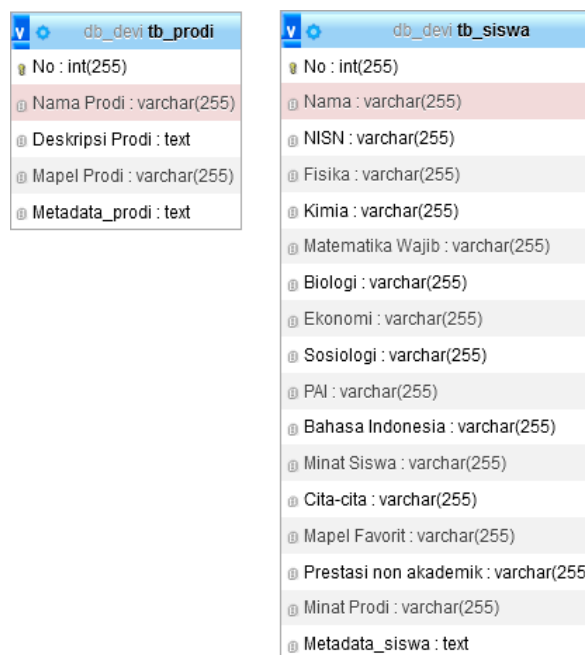
Analisis dalam bab ini akan menguraikan dengan rinci bagaimana kedua data tersebut diolah untuk memberikan wawasan mengenai keterkaitan minatsiswa dengan pilihan program studi yang relevan. Proses pengolahan data meliputi langkah-langkah seperti pembobotan kata, penghitungan kemiripan, serta 35. Penyajian hasil dalam bentuk matriks. Hasil dari pengolahan data tersebut akan dibahas secara mendalam, termasuk interpretasi terhadap keterkaitan yang ditemukan, serta implikasi bagi siswa dalam menentukan pilihan studi mereka. Untuk informasi yang lebih lengkap dan komprehensif, pembaca dapat merujuk pada lampiran 1, yang menyajikan data dan analisis tambahan.

4.2 Implementasi Data.

Implementasi basis data merupakan gambaran dari proses pembuatan dan pengelolaan database yang mendukung sistem yang dikembangkan. Database ini dibangun menggunakan MySQL untuk memastikan fungsionalitas sistem berjalan dengan optimal dan dapat diakses dengan mudah. Dalam konteks penelitian ini, hanya digunakan dua tabel utama, yaitu tabel "tb_siswa" dan tabel "tb_prodi," yang masing-masing memiliki fungsi yang sangat penting. Tabel "tb_siswa" berisidata mentah mengenai siswa, termasuk informasi pribadi dan minat mereka sebelum melalui tahap prapemrosesan. Data dalam tabel ini mencakup berbagai atribut, seperti nama usia, dan data relevan lainnya yang akan diolah lebih lanjut. Tabel ini berperan sebagai sumber informasi dasar yang menjadi titik awal analisis, sehingga informasi yang ada di dalamnya sangat krusial bagi penelitian. Sementara itu, tabel "tb_prodi" menyimpan informasi mengenai program studi yang telah melalui proses

prapemrosesan. Proses ini meliputi pembersihan data, pengelompokan, dan pengolahan lainnya yang diperlukan untuk meningkatkan kualitas informasi yang disimpan. Tabel ini berfungsi untuk mencocokkan minat siswa dengan program studi yang tersedia, dan memberikan data yang siap untuk dianalisis dalam konteks rekomendasi pendidikan

Kedua tabel ini merupakan bagian integral dari sistem yang dikembangkan, dan implementasi serta strukturnya dijelaskan secara rinci pada Gambar 4.1. Untuk memastikan integrasi yang baik antara sistem dan database, pembuatan database dilakukan dengan menggunakan MySQL yang dikonfigurasi dengan library from flask_mysql import MySQL. Konfigurasi ini penting untuk menghubungkan aplikasi berbasis Flask dengan database MySQL, memungkinkan akses data yang efisien dan pengolahan yang cepat. Dengan demikian, implementasi basis data dalam penelitian ini tidak hanya mendukung fungsionalitas sistem, tetapi juga memberikan fondasi yang kuat untuk analisis data yang lebih mendalam.



db_devi tb_prodi	
No	: int(255)
Nama Prodi	: varchar(255)
Deskripsi Prodi	: text
Mapel Prodi	: varchar(255)
Metadata_prodi	: text

db_devi tb_siswa	
No	: int(255)
Nama	: varchar(255)
NISN	: varchar(255)
Fisika	: varchar(255)
Kimia	: varchar(255)
Matematika Wajib	: varchar(255)
Biologi	: varchar(255)
Ekonomi	: varchar(255)
Sosiologi	: varchar(255)
PAI	: varchar(255)
Bahasa Indonesia	: varchar(255)
Minat Siswa	: varchar(255)
Cita-cita	: varchar(255)
Mapel Favorit	: varchar(255)
Prestasi non akademik	: varchar(255)
Minat Prodi	: varchar(255)
Metadata_siswa	: text

Gambar 4.1 Database Sistem

Tabel "tb_prodi" terdiri dari beberapa kolom yang mencakup informasi penting tentang program studi. Kolom pertama adalah No, yang berfungsi sebagai identifier unik untuk setiap program studi, sehingga memudahkan dalam pengelolaan dan pencarian data. Kolom Nama Prodi menyimpan nama lengkap dari program studi, yang menjadi referensi utama dalam analisis dan rekomendasi. Selanjutnya, kolom Deskripsi Prodi memberikan penjelasan mendetail mengenai karakteristik dan tujuan dari program studi tersebut, sehingga dapat membantu siswa dalam memahami pilihan yang tersedia. Kolom Mapel Prodi mencantumkan mata pelajaran yang relevan dengan program studi, yang penting untuk menilai kesesuaian minat siswa. Terakhir, kolom Metadata_prodi merupakan gabungan dari semua kolom di tabel ini.

Sementara itu, tabel "tb_siswa" berisi data siswa dengan berbagai kolom yang dirancang untuk mencatat informasi menyeluruh mengenai siswa. Kolom pertama adalah No, yang berfungsi sebagai identifier untuk setiap siswa, mirip dengan fungsinya di tabel prodi. Kolom Nama digunakan untuk menyimpan nama lengkap siswa, sedangkan kolom NISN mencatat nomor induk siswa yang bersifat unik dan penting untuk identifikasi. Selain itu, tabel ini mencakup kolom-kolom yang berisi nilai dari berbagai mata pelajaran, seperti Fisika, Kimia, dan Biologi, yang mencerminkan prestasi akademik siswa. Tabel ini juga mencakup kolom-kolom seperti Minat Siswa, Cita-cita, Mapel Favorit, Prestasi Non-Akademik, dan Minat Prodi, yang merekam minat serta preferensi siswa terkait dengan bidang akademik maupun non-akademik. Informasi ini sangat berharga dalam merumuskan rekomendasi program studi yang sesuai dengan bakat dan aspirasi

siswa. Kolom Metadata_siswa berfungsi sebagai gabungan dari beberapa kolom penting seperti minat siswa, cita-cita, mapel favorit, dan prestasi, memberikan gambaran holistik mengenai profil siswa. Dengan demikian, struktur tabel ini dirancang untuk memudahkan analisis dan pengolahan data, serta mendukung tujuan utama penelitian dalam mencocokkan minat siswa dengan program studi yang relevan.

```
# KONFIGURASI DATABASE
app.config["MYSQL_HOST"] = "localhost"
app.config["MYSQL_USER"] = "root"
app.config["MYSQL_PASSWORD"] = ""
app.config["MYSQL_DB"] = "db_devi"
mysql = MySQL(app)
```

Gambar 4.2 Konfigurasi Database

Konfigurasi database di atas menggunakan library Flask-MySQLdb untuk menghubungkan aplikasi Flask dengan database MySQL. Library ini dirancang untuk memberikan kemudahan dalam integrasi antara aplikasi berbasis Flask dan database MySQL, sehingga memudahkan pengelolaan data dan interaksi dengan database secara efisien. Dalam konfigurasi ini, parameter MYSQL_HOST disetel ke "localhost," yang menunjukkan bahwa database dijalankan di komputer lokal, menjadikannya mudah diakses selama pengembangan dan pengujian. Selanjutnya, MYSQL_USER diatur ke "root," yang merupakan pengguna default MySQL. Penggunaan pengguna root ini umumnya dilakukan dalam lingkungan pengembangan untuk memberikan akses penuh ke database, meskipun dalam praktik produksi, disarankan untuk membuat pengguna dengan hak akses terbatas demi keamanan. Untuk parameter MYSQL_PASSWORD, dalam konfigurasi ini

tidak diisi (kosong), yang mengindikasikan bahwa pengguna root diatur tanpa password, sebuah pengaturan yang mungkin cocok untuk lingkungan pengembangan tetapi perlu ditangani dengan hati-hati di lingkungan produksi untuk menghindari potensi masalah keamanan.

4.3 Preprocessing

Pada tahap prapemrosesan data, dilakukan beberapa langkah seperti pembersihan data, *case folding* untuk mengubah teks menjadi huruf besar, penghapusan kata umum (*stopword removal*), serta stemming untuk mengembalikan kata ke bentuk dasarnya. Keempat langkah ini memanfaatkan beberapa *library*, yang prosesnya digambarkan dalam kode program berikut:

```

from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

def preprocessing_input(datisiswa):
    # Cleaning: Menghapus karakter non-alfanumerik dan angka
    datisiswa = ''.join([char for char in datisiswa if char.isalnum() or char.isspace()])

    # Case Folding: Mengubah teks menjadi lowercase
    datisiswa = datisiswa.lower()

    # Stopword Removal
    factory = StopWordRemoverFactory()
    stopwords_removal = factory.create_stop_word_remover()
    datisiswa = stopwords_removal.remove(datisiswa)

    # Stemming
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    datisiswa = stemmer.stem(datisiswa)

    return datisiswa

```

Gambar 4.3 Code Preprocessing

Pada tahap *cleaning data*, proses ini dilakukan untuk membersihkan teks dari karakter-karakter yang tidak diinginkan, seperti tanda baca atau simbol yang tidak relevan dalam analisis teks. Kode "[char for char in datisiswa if char.isalnum() or char.isspace()]" digunakan untuk mengiterasi setiap karakter(char) dalam

string data siswa. Selanjutnya, mengecek apakah karakter tersebut adalah huruf, angka, atau spasi. Jika kondisinya terpenuhi, karakter tersebut akan tetap ada (bernilai true), sedangkan tanda baca akan dihapus dan digantikan dengan karakter kosong ')'. Hal ini memastikan bahwa hanya karakter yang diperlukan yang tersisa dalam teks. Contoh hasil dari implementasi manual

Sebelum: "Data siswa ini sangat penting dalam proses analisis! Mereka berfokus untuk belajar lebih giat."

Sesudah: "Data siswa ini sangat penting dalam proses analisis Mereka berfokus untuk belajar lebih giat".

Setelah pembersihan selesai, teks diubah menjadi huruf kecil dengan menggunakan `lower()`. Langkah ini penting untuk memastikan konsistensi dalam analisis teks, karena dengan mengubah semua huruf menjadi kecil, perbedaan antara huruf besar dan kecil tidak akan memengaruhi proses analisis. Misalnya, kata "Siswa" dan "siswa" akan dianggap sama setelah proses ini, sehingga membantu mencegah adanya ketidakcocokan dalam perbandingan kata.

Contoh hasil implementasi manual :

Sebelum: "Data siswa ini sangat penting dalam proses analisis"

Sesudah: "data siswa ini sangat penting dalam proses analisis"

Langkah berikutnya adalah *stopword removal*, yang dilakukan dengan menggunakan *Sastrawi*. Dalam tahap ini, `StopWordRemoverFactory` digunakan untuk membuat objek *stopword remover*. Setelah objek tersebut dibuat, teks kemudian diproses menggunakan `remove()`, yang berfungsi menghapus kata-kata umum yang tidak relevan dalam analisis, seperti kata sambung atau kata-kata yang

tidak memiliki makna penting dalam konteks teks.

Contoh hasil implementasi manual :

Sebelum: "Data siswa ini sangat penting dalam proses analisis"

Sesudah: "data siswa penting dalam proses analisis"

Pada tahap akhir, dilakukan stemming dengan memanfaatkan *Sastrawi* melalui *StemmerFactory* untuk membuat objek *stemmer*. *Stemming* adalah proses penting yang mengubah kata-kata dalam teks ke bentuk dasarnya, seperti mengubah "berlari" menjadi "lari". Hal ini bertujuan untuk mengurangi variasi kata dan menyederhanakan analisis. Setelah semua proses prapemrosesan ini selesai, teks siap untuk dianalisis lebih lanjut.

Contoh hasil implementasi manual :

Sebelum: "data siswa sedang belajar berbagai teknik komputer"

Sesudah: "data siswa sedang ajar bagai teknik komputer"

Tabel 4.1 Tabel Hasil Preprocessing

No	Nama	NISN	MinatSiswa	Cita-cita	Mapel Favorit	MinatProdi
1	AbdulRazak	69052714	Seni Ukir Tenis Meja	Usaha	Pendidikan Agama Islam	Biologi
2	Alfi Karimatil Ma'rifah	133689163	Baca & Nyanyi	Guru agama/hafid zah	Pendidikan Agama Islam	Manajemen
3	Moh Fiqri Arrasid	75167933	Gambar, Menulis Fiksi	Rencana wilayah kota sekaligus tuliskanovel	Fisika, Biologi, Bahasa Inggris	Arsitektur
4	Nurul Izzati	75131889	Nyanyi	Bidan	Kimia	Teknik Informatika
5	Silvia Rossiana Sandra	79105432	Menulis Novel	Dokter psikologi	Biologi, Bahasa Inggris	Hubungan Internasio nal

No	Nama	NISN	MinatSiswa	Cita-cita	Mapel Favorit	MinatProdi
6	Siti Nur Aini	3094981170	Menulis, Main Voli	Dokter	Ekonomi	Akuntansi
7	Anis Safara	75807768	Masak	Perawat	Biologi	Biologi
8	Nur Aini	3098268056	Baca	Dokter	Bahasa Indonesia	Biologi

4.4 Metadata

Setelah tahap prapemrosesan selesai, yang mencakup proses cleaning hingga stemming, langkah selanjutnya adalah pembuatan kolom baru yang dinamakan "metadata." Kolom ini dirancang untuk menyimpan informasi tambahan yang relevan dan memberikan konteks yang lebih dalam untuk analisis data siswa. Penggunaan kolom metadata bertujuan untuk mengintegrasikan berbagai aspek yang mempengaruhi pilihan dan minat siswa dalam pendidikan

Dalam kolom metadata ini, terdapat beberapa sub-kolom penting yang masing-masing berfungsi untuk mencatat informasi spesifik mengenai siswa. Pertama, sub-kolom 'Minat Siswa' mencerminkan minat akademik dan non-akademik setiap siswa, memberikan gambaran tentang bidang-bidang yang menarik bagi mereka. Kedua, sub-kolom 'Cita-cita' berisi aspirasi atau tujuan karier yang ingin dicapai oleh siswa, yang dapat mencerminkan harapan dan motivasi mereka dalam memilih program studi.

Selanjutnya, ada sub-kolom 'Mapel Favorit,' yang menunjukkan mata pelajaran yang paling disukai oleh siswa, yang bisa menjadi indikator mengenai kekuatan akademis mereka. Sub-kolom 'Prestasi Non-Akademik' mencatat pencapaian siswa

di luar bidang akademik, seperti kejuaraan olahraga atau keterlibatan dalam kegiatan ekstrakurikuler, yang juga berkontribusi terhadap pengembangan karakter dan keterampilan sosial siswa. Terakhir, sub-kolom 'MinatProdi' menggambarkan ketertarikan siswa terhadap program studi tertentu, yang sangat penting dalam konteks rekomendasi pendidikan. Dengan adanya kolom metadata ini, diharapkan analisis data dapat memberikan wawasan yang lebih mendalam mengenai karakteristik dan preferensi siswa. Hal ini tidak hanya membantu dalam memahami pola minat dan bakat siswa, tetapi juga berperan penting dalam merumuskan rekomendasi yang lebih tepat dan relevan sesuai dengan kebutuhan dan aspirasi mereka. Dengan demikian, informasi yang terkumpul melalui kolom metadata ini diharapkan dapat memberikan nilai tambah dalam proses pengambilan keputusan mengenai pilihan program studi yang sesuai bagi setiap siswa.

Tabel 4.2 Tabel Hasil Metadata

No	Nama	NISN	Minat Siswa	Cita-cita	Mapel Favorit	Minat Prodi	Metadata
1	Abdul Razak	69052714	Seni Ukir Tenis Meja	Usaha	Pendidikan Agama Islam	Biologi	Seni ukir tenis meja usaha pai olahraga mungkin biologi
2	Alfi Karimatil Ma'rifah	133689163	Baca & Nyanyi	Guru agama/hafidzah	Pendidikan Agama Islam	Manajemen	Baca nyanyi guru Agama hafi dzah pai senin musik manajemen

No	Nama	NISN	Minat Siswa	Cita-cita	Mapel Favorit	Minat Prodi	Metadata
3	Moh Fiqri Arrasid	75167933	Gambar, Menulis Fiksi	Rencana wilayah kota sekaligus tulis novel	Fisika, Biologi, Bahasa Inggris	Arsitektur	Gambar tulis fiksi rencana wilayah kota sekaligus tulis novel fisika biologi bahasa inggris ketua sebid 7 osis arsitektur
4	Nurul Izzati	75131889	Nyanyi	Bidan	Kimia	Teknik Informatika	Nyanyi bidan kimia musik teknik informatika
5	Silvy Rossiana Sandra	79105432	Menulis Novel	Dokter psikologi	Biologi, Bahasa Inggris	Hubungan Internasional	Tulis novel dokter psikologi biologi bahasa inggris ada hubungan internasional
6	Siti NurAini	3094981170	Menulis, Main Voli	Dokter	Ekonomi	Akuntansi	Menulisma in voly dokter ekonomi sertifikat ahli khusus akuntansi
7	Anis Safara	75807768	Masak	Perawat	Biologi	Biologi	Masak perawat biologi ada biologi
8	NurAini	3098268056	Baca	Dokter	Bahasa Indonesia	Biologi	

4.5 Perhitungan TF-IDF

Setelah data melalui proses preprocessing dan digabungkan dalam kolom metadata, langkah selanjutnya adalah melakukan pembobotan kata dengan mengonversinya ke dalam bentuk matriks menggunakan algoritma TF-IDF (Term Frequency-Inverse Document Frequency). Proses transformasi data menjadi matriks ini penting untuk mengukur relevansi kata-kata dalam dokumen dibandingkan dengan seluruh koleksi dokumen yang ada. Dengan menggunakan algoritma TF-IDF, kita dapat menekankan kata-kata yang lebih sering muncul dalam dokumen tertentu, namun jarang ditemukan di dokumen lainnya, sehingga dapat memberikan bobot yang lebih besar pada kata-kata yang lebih informatif. Proses ini dapat dilakukan dengan menggunakan fungsi-fungsi yang tersedia dalam `TfidfVectorizer`, yang merupakan alat utama dalam implementasi algoritma TF-IDF. `TfidfVectorizer` secara otomatis menangani berbagai tahapan, mulai dari tokenisasi hingga penghitungan frekuensi kata dan penentuan bobot.

Gambar 4.4 di bawah ini menunjukkan cara kerja dari fungsi-fungsi ini serta bagaimana data yang telah diproses dapat diubah menjadi matriks TF-IDF. Dengan matriks ini, kita akan memiliki representasi numerik yang dapat digunakan untuk analisis lebih lanjut, termasuk penghitungan kemiripan antarkonten menggunakan metode yang telah dibahas sebelumnya. Transformasi ini merupakan langkah krusial dalam rangka mempersiapkan data untuk model rekomendasi yang lebih efektif, seperti yang ditunjukkan pada Gambar 4.4.

```
# Calculate TF-IDF and cosine similarity
vectorizer = TfidfVectorizer()
tfidf_matrix = vectorizer.fit_transform([input_text_cleaned] + [p['Metadata_prodi'] for p in prodi])
tfidf_matrix = normalize(tfidf_matrix)
```

Gambar 4.4 Code TFDF

Proses dimulai dengan menginisialisasi objek `TfidfVectorizer()`, yang berfungsi untuk mengonversi teks menjadi representasi numerik berdasarkan metode TF-IDF (Term Frequency-Inverse Document Frequency). Metode TF-IDF adalah teknik yang umum digunakan dalam pemrosesan teks untuk mengevaluasi pentingnya kata dalam suatu dokumen dalam konteks seluruh kumpulandokumen. Proses ini menghitung frekuensi kemunculan kata dalam dokumen tertentu, serta seberapa jarang kata tersebut muncul di seluruh koleksi dokumen. Dengan demikian, kata-kata yang sering muncul dalam satu dokumen tetapi jarang muncul di dokumen lain akan memiliki bobot yang lebih tinggi. Sebaliknya, kata-kata umum yang sering muncul di banyak dokumen akan memiliki bobot yang lebih rendah. Ini memungkinkan model untuk lebih menekankan kata-kata yang lebih unik dan relevan dalam konteks analisis.

Setelah proses inisialisasi selesai, metode `fit_transform` diterapkan pada kumpulan teks yang terdiri dari teks input yang telah dibersihkan, yang disebut sebagai `input_text_cleaned`, dan metadata dari objek program studi. Langkah ini menghasilkan matriks TF-IDF yang merepresentasikan hubungan antar kata dalam teks, di mana setiap baris mewakili dokumen dan setiap kolom mewakili kata unik. Selanjutnya, matriks tersebut dinormalisasi menggunakan fungsi `normalize()`. Proses normalisasi ini bertujuan untuk memastikan bahwa setiapvektor memiliki panjang yang sama, sehingga membantu menjaga konsistensi dalam algoritma yang akan digunakan selanjutnya. Hasil akhirnya adalah matriks numerik yang siap digunakan untuk analisis lanjutan, seperti yang ditunjukkan pada Tabel 4.2. Matriks ini tidak hanya menjadi dasar bagi berbagai analisis, tetapi juga berperan penting

dalam proses penghitungan kemiripan konten yang akan dilakukan selanjutnya. seperti pada tabel 4.2.

Tabel 4.3 Hasil TfIdf Term

No	Term	Tf	Idf	W
1	abdi	0,138	3,4849	0,6188
2	adab	0,2472	3,0794	1,0085
3	adalah	0,1337	3,4849	0,5994
..	...			
368	utama	0,1282	3,4849	0,5751
369	visual	0,3924	3,0794	1,6009
370	yang	0,2102	3,0794	0,8576

4.6 Cosine Similarity

Tahap selanjutnya, setelah mendapatkan pembobotan kata, adalah melakukan perhitungan kemiripan antar konten atau data dengan menggunakan algoritma Cosine Similarity. Proses ini penting untuk mengevaluasi seberapa mirip dua dokumen atau konten berdasarkan vektor pembobotan yang telah ditentukan sebelumnya. Dalam contoh berikut, kita akan mengaplikasikan algoritma ini pada data program studi (prodi) untuk mengukur kesamaan antara berbagai program yang tersedia. Penerapan algoritma ini akan memberikan wawasan yang lebih dalam mengenai kecocokan antara minat siswa dan program studi yang ada, serta membantu dalam memberikan rekomendasi yang lebih akurat. Perhitungan akan dilakukan dengan menggunakan rumus pada persamaan 2.4 dan hasil perhitungannya ditunjukkan pada tabel 3.9.

Untuk menghitung cosine similarity maka akan dilakukan perhitungan dot product dan panjang vektor terlebih dahulu. Dalam dokumen dibawah ini profil

siswa dan prodi D1,D2 dan D3, dilakukan sebuah perkalian nilai TF- 28 IDF dari setiap kata kunci dalam siswa dengan nilai yang sesuai dari prodi, kemudian hasilnya dijumlahkan.

Dari tabel dibawah ini kita akan melakukan perhitungan antara dokumen Dq dan Dq, dokumen Dq dan D1, dokumen Dq dan D2, dokumen Dq dan D3. Dibawah ini contoh perhitungan dot Product pada Dq D1

$$\begin{aligned}
 \text{Dot Product} &= (0 \times 0) + (0 \times 0) + (0 \times 0) + (0 \times 0) + (0 \times 0) + (0 \times 0, 60067) + \\
 &(0 \times 0) + (0 \times 0) + (0 \times 0) + (0 \times 0) + (0 \times 0) + (0 \times 0) + (0 \times 0) + (0 \times 0) + (0 \times \\
 &0) = 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0
 \end{aligned}$$

Nilai diatas adalah hasil dari perhitungan antara TF-Dq – TF-D3 dengan cara mengalikan disetiap kata kunci lalu hasilnya dijumlahkan

Tabel 4.4 Hasil Perhitungan Dot Product

TERM	A.B			
	Dq Dq	Dq D1	Dq D2	Dq D3
abdi	0	0	0	0
adab	0	0	0	0
adalah	0	0	0	0
administrasi	0	0	0	0
ahli	0	0	0	0

ajar	0	0	0	0
akademik	0	0	0	0
akademis	0	0	0	0
akan	0	0	0	0
aktif	0	0	0	0
aktivis	0	0	0	0
aku	0	0	0	0
akuntan	0	0	0	0
akutansi	0	0	0	0
Hasil Dot Product	0	0	0	0

Hasil perhitungan diatas menunjukkan bahwa vektor query dan vektor dokumen tidak memiliki kesamaan term atau nilai bobot term yang dihitung semuanya adalah nol.

Query dan dokumen kosong (tidak memiliki kesamaan) Proses pembobotan seperti TF atau TF-IDF menghasilkan nilai nol untuk semua term.

Term yang digunakan tidak muncul sama sekali dalam dokumen maupun query.

Langkah kedua adalah perhitungan panjang dari vektor TF-IDF siswa (D_q) dan $prodi/D1, D2, D3$, yaitu dengan menghitung akar kuadrat dari jumlah kuadrat semua nilai dalam vektor. berikut ini hasil perhitungannya yang ditunjukkan pada tabel 4.6.

$$A = \sqrt{0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2} = \sqrt{0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0} = 0$$

Dari proses perhitungan diatas ditemukan hasil panjang vector dari D_q adalah **0** dan seterusnya.

Tabel 4.5 Hasil Perhitungan Panjang Vektor

TERM	A.B			
	Dq Dq	Dq D1	Dq D2	Dq D3
Abdi	0	0	0	0
Adab	0	0	0	0
adalah	0	0	0	0
administrasi	0	0	0	0
ahli	0	0	0	0
ajar	0	0,0035	0,0009	0,0024
akademik	0	0	0	0
akademis	0	0	0	0,0212
akan	0	0	0	0
aktif	0	0	0	0
aktivis	0	0	0	0
aku	0	0	0	0
akuntan	0	0	0	0
akutansi	0	0	0	0
Hasil Panjang Vektor	0	0,0587	0,03	0,1536

Mayoritas nilai dalam tabel adalah 0, menunjukkan bahwa banyak istilah (term) tidak terdapat atau tidak berkontribusi pada perhitungan dot product. Namun ada beberapa nilai positif kecil seperti :

Pada “ajar” di **Dq D1**, **Dq D2**, dan **Dq D3**.

Pada “akademis” di **Dq D3**

Jadi nilai di atas mengindikasikan keberadaan kata “ajar” dan “akademis” dalam dokumen tertentu serta kontribusinya dalam perhitungan, hasil panjang vektor di bagian akhir adalah perhitungan panjang atau magnitudo dari vektor query dan dokumen, panjang vektor menunjukkan seberapa besar kontribusi dari setiap dokumen terhadap query yang diberikan:

Dq Dq: 0 (query memiliki nilai nol)

Dq D1: 0,587.

Dq D2: 0,03.

Dq D3: 0,1536

Panjang vektor (magnitudo) digunakan untuk menghitung kemiripan antara query dan dokumen-dokumen (D1, D2, D3). Semakin tinggi nilai panjang vektor, semakin relevan dokumen tersebut terhadap query. D3 (0,1536) memiliki nilai tertinggi dibandingkan dengan D1 (0,0587) dan D2 (0,03). Ini menunjukkan bahwa dokumen D3 lebih relevan terhadap query dibandingkan dokumen lainnya

Dari proses perhitungan yang telah dilakukan dengan TF-IDF dot product dan panjang vektor. Langkah selanjutnya yaitu menghitung cosine similarity, dengan rumus dibawah ini :

$$\text{cosine similarity} = \frac{A \cdot B}{\|A\| \|B\|}$$

Keterangan :

- A . B adalah hasil perkalian dot (produk) antara vektor A dan B.
- ||A|| adalah panjang vektor A
- ||B|| adalah panjang vektor

$$\text{Cosine Similarity (Dq, Dq)} = \frac{0}{0 \times 0} = \frac{0}{0} = 0$$

$$\text{Cosine Similarity (Dq, D1)} = \frac{0}{0,0587 \times 0,0587} = \frac{0}{0,0034} = 0$$

$$\text{Cosine Similarity (Dq, D2)} = \frac{0}{0,03 \times 0,03} = \frac{0}{0,0009} = 0$$

$$\text{Cosine Similarity (Dq, D3)} = \frac{0}{0,1536 \times 0,1536} = \frac{0}{0,0235} = 0$$

Di bawah ini, Anda dapat melihat implementasi dari proses ini dalam bentuk source code:

```
def calculate_cosine_similarity(data):
    vectorizer = TfidfVectorizer()
    tfidf_matrix = vectorizer.fit_transform(data['Metadata_prodi'])

    # Hitung cosine similarity dari matriks TF-IDF
    cosine_sim = cosine_similarity(tfidf_matrix, tfidf_matrix)

    cosine_sim_df = pd.DataFrame(cosine_sim, columns=data.index)

    return cosine_sim_df
```

Gambar 4.5 Code *Cosine Similarity*

Pada kode program tersebut, fungsi `calculate_cosine_similarity` digunakan untuk menghitung kemiripan antar teks dengan memanfaatkan metode cosine similarity. Proses ini dimulai dengan menginisialisasi objek `TfidfVectorizer()`, yang berfungsi untuk mengubah teks dalam kolom `Metadata_prodi` menjadi matriks TF-IDF. Matriks ini merepresentasikan kata-kata dalam teks dengan bobot yang dihitung berdasarkan frekuensi kemunculannya dalam dokumen tertentu serta seberapa jarang kata tersebut muncul di dokumen lainnya. Konsep ini membantu

dalam menekankan kata-kata yang lebih unik dan relevan, sehingga dapat meningkatkan kualitas analisis yang dilakukan.

Setelah matriks TF-IDF terbentuk, fungsi `cosine_similarity()` diterapkan untuk menghitung derajat kemiripan kosinus antara setiap pasangan dokumen dalam dataset. Dua dokumen dianggap mirip jika nilai cosine similarity mereka mendekati angka 1, yang menandakan bahwa kedua dokumen memiliki konten yang sangat mirip. Sebaliknya, nilai yang mendekati 0 menunjukkan bahwa dokumen-dokumen tersebut memiliki sedikit atau tidak ada kesamaan. Proses ini sangat penting dalam konteks rekomendasi, di mana pemahaman tentang kemiripan antar program studi dapat membantu siswa dalam menentukan pilihan yang sesuai dengan minat dan bakat mereka. Dengan demikian, algoritma ini tidak hanya mengoptimalkan pencocokan konten, tetapi juga meningkatkan pengalaman pengguna dalam memilih program studi yang tepat. Seperti pada tabel dibawah ini.

Tabel 4.6 Hasil Perhitungan Cosine Similarity

Dq Dq	Dq D1	Dq D2	Dq D3
0	0	0	0

4.7 Implementasi Sistem

Sistem rekomendasi dalam penelitian ini dikembangkan dalam bentuk website menggunakan framework Flask. Flask adalah framework yang memungkinkan integrasi antara Python untuk pengolahan data dan HTML, CSS, serta JavaScript untuk tampilan dan visualisasi data. Berikut adalah beberapa contoh tampilan dari sistem yang telah dibangun.

Implementasi system rekomendasi dalam penelitian ini dilakukan melalui

pengembangan aplikasi berbasis web yang menggunakan framework Flask. Flask dipilih karena kesederhanaannya dalam mendukung integrasi antara Python sebagai Bahasa pemrograman utama untuk pengolahan data dan, HTML, CSS, serta JavaScript untuk bagian antarmuka dan visualisasi data. Dengan memungkinkan peneliti untuk menghubungkan logika rekomendasi dengan tampilan antarmuka secara efektif dan efisien.

Aplikasi ini dirancang untuk memberikan pengalaman pengguna yang optimal dan memperoleh rekomendasi program studi yang relevan berdasarkan minat siswa. Dalam prosesnya, data dari siswa dianalisis dan kemudian disajikan dalam bentuk rekomendasi melalui antar muka yang mudah diakses oleh pengguna. Setiap elemen dari tampilan, mulai dari halaman utama, hasil rekomendasi, hingga fitur tambahan lainnya, diatur secara optimal agar dapat memberikan interaksi yang intuitif.

Dengan mengombinasikan Python pada sisi server dan HTML/CSS serta JavaScript pada sisi klien, system ini tidak hanya menyediakan hasil rekomendasi yang akurat, tetapi juga menghadirkan visualisasi Data yang jelas dan menarik bagi pengguna. Pengguna dapat langsung melihat hasil rekomendasi melalui antarmuka yang responsive, yang menyesuaikan diri dengan perangkat yang digunakan baik itu desktop, tablet, atau smartphone.

4.7.1 Landing Page

Pada halaman landing page terdapat berbagai elemen yang dirancang untuk memberikan kemudahan navigasi dan pengalaman pengguna yang informatif. Salah satu elemen utama pada halaman ini adalah Navbar, yang berfungsi sebagai menu

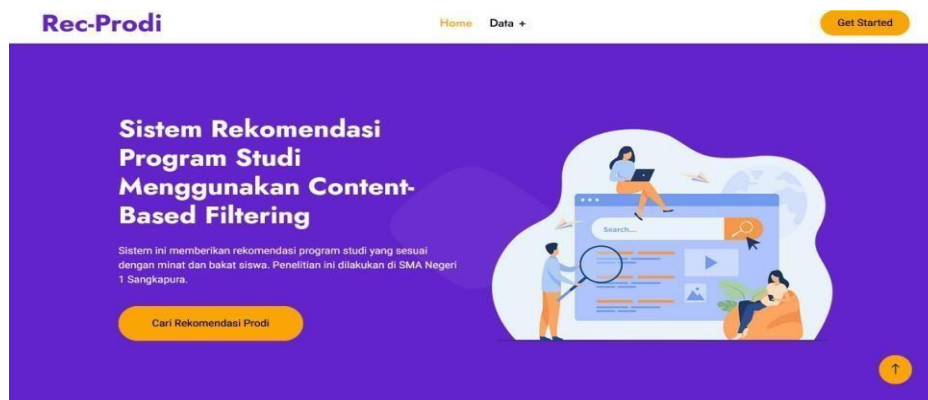
navigasi utama. Navbar ini mencakup beberapa pilihan, seperti “Data minat siswa” dan “data program studi”, yang memungkinkan pengguna untuk dengan mudah berpindah antara halaman yang berisi data terkait minat siswa dan informasi tentang program studi tersedia. Dengan adanya Navbar, pengguna dapat mengakses informasi penting hanya dengan satu klik, sehingga mempermudah proses pencarian dan pemilihan program studi yang sesuai dengan minat mereka.

Selain Navbar, halaman landing page juga dilengkapi dengan welcoming page atau halaman sambutan. Bagian ini menampilkan judul website dan deskripsi singkat mengenai fungsi dan tujuan website. Pada bagian judul, tercantum nama website, misalnya “ Sistem Rekomendasi Program Studi Berbasis Minat Siswa.” Judul ini disusun dengan gaya yang menarik dan mudah terbaca untuk memberikan kesan professional dan ramah kepada pengguna baru.

Deskripsi singkat pada welcoming page ini menjelaskan tujuan dari system ini, yakni membantu siswa dalam menemukan program studi yang sesuai dengan minat mereka melalui metode Content-based filtering. Penjelasan ini memberikan gambaran singkat tentang bagaimana system bekerja Dan bagaimana siswa dapat memanfaatkannya untuk memilih jurusan yang paling sesuai. Selain itu, deskripsi ini dapat menyertakan kalimat motivasi atau ajakan yang mendorong siswa untuk segera mencoba fitur rekomendasi yang tersedia.

Visual yang menarik, seperti warna yang harmonis dan tata letak yang intuitif, juga menjadi bagian penting dari landing page ini. Elemen-elemen visual seperti ikon, Ilustrasi, atau gambar latar belakang yang relevan dengan dunia Pendidikan dapat di tambahkan untuk menambah daya Tarik visual sekaligus

meningkatkan pemahaman pengguna.



Gambar 4.6 Landing Page

4.7.2 Data Siswa Page

Pada halaman data Siswa Page, pengguna akan melihat seluruh data siswa yang telah terdaftar atau terinput ke dalam database system dalam bentuk table yang terstruktur. Table ini dirancang untuk memudahkan pengguna dalam membaca dan memahami informasi yang telah dikumpulkan. Data yang ditampilkan di halaman ini meliputi hasil preprocessing dari data yang dikumpulkan melalui Google-Form, di mana siswa telah mengisi informasi dasar mengenai minat akademik, cita-cita, serta nilai mata pelajaran yang mereka kuasai. Data ini kemudian diekspor ke dalam database system untuk memudahkan proses pemanggilan dan penampilan data.

Tabel pada Data Siswa Page terdiri dari beberapa kolom yang mencakup berbagai informasi penting, seperti Nama Siswa, Nomor Induk Siswa Nasional (NISN), Minat Akademik, Mata Pelajaran Favorit, Prestasi Non-Akademik, dan Program Studi yang diminati. Setiap kolom dirancang agar pengguna dapat melihat data siswa secara lebih menyeluruh dan mendalam. Data tersebut telah melalui proses preprocessing yang meliputi penghapusan data duplikat,

normalisasi, dan konversi teks agar konsisten, sehingga lebih siap untuk dianalisis dan digunakan dalam system rkeomendasi.

Selain itu, pada halaman ini juga terdapat fungsi-fungsi tambahan yang emudahkan pengguna dalam melakukan filter dan sortir data. Misalnya, pengguna dapat melakukan pencarian dengan memasukkan nama atau kata kunci tertentu untuk menemukan data siswa secara cepat. Fitur sortir juga memungkinkan pengguna untuk mengurutkan data bedasarkan kolom teretntu, seperti mengurutkan berdasarkan nama atau mata pelajaran favorit siswa. Fitur ini memudahkan dalam proses analisis data, khususnya bagi admin atau guru yang ingin melihat profil minat siswa secara spesifik.

Selain fitur pencarian dan pengurutan, table data siswa juga memungkinkan pengguna untuk mengekspor data dalam berbagai format, seperti PDF atau Exel, sehingga mempermudah pemanfaatan data dalam kegiatan lain, seperti pelaporan atau evaluasi Pendidikan. Pada bagian bawah halaman ini terdapat tombol untuk menambahkan data siswa baru atau memperbarui data yang telah ada jika diperlukan. seperti pada Gambar 4.7.

No	Nama	NISN	Minat Siswa	Cita-Cita	Mapel Favorit	Minat Prodi	Metadata	
1	Abdul Razak	69052714	seni ukir tenis meja	usaha	pai	biologi	seni ukir tenis meja usaha pai olahraga mungkin biologi	Cek Rekomendasi Prodi
2	Alfi karimati ma'rifah	133689163	baca nyanyi	guru agamahafidzah	pai	manajemen	baca nyanyi guru agamahafidzah pai seni musik manajemen	Cek Rekomendasi Prodi
3	Moh figri arrasid	75167933	gambar tulis fiksi	rencana wilayah kota sekaligus tulis novel	fisika biologi bahasa inggris	arsitektur	gambar tulis fiksi rencana wilayah kota sekaligus tulis novel fisika biologi bahasa inggris ketua sekbid 7 osis arsitektur	Cek Rekomendasi Prodi
4	Nurul Izzati	75131889	nyanyi	bidan	kimia	teknik informatika	nyanyi bidan kimia musik teknik	Cek Rekomendasi Prodi

Gambar 4.7 Data Page

Pada halaman ini, setiap siswa yang telah mengisi form minat bakat akan memiliki informasi yang terstruktur dalam beberapa kolom. Kolom-kolom ini berisi data yang relevan seperti Nama Siswa, NISN, Minat Akademik. Data tersebut membantu dalam memberikan gambaran tentang minat bakat masing-masing siswa, yang digunakan sebagai dasar untuk merekomendasikan program studi yang sesuai.

Di samping setiap entri siswa, terdapat tombol “Cek Rekomendasi” yang memungkinkan pengguna untuk langsung melihat rekomendasi program studi. Saat tombol ini di klik, system akan mengarahkan pengguna ke halaman rekomendasi prodi. Halaman ini akan menampilkan daftar program studi yang sesuai dengan profil minat bakat siswa yang telah dipilih. Rekomendasi ini disusun menggunakan metode content-based filtering, yang memadukan informasi minat dan prestasi siswa dengan atribut dari berbagai program studi untuk menemukan program yang paling sesuai.

Pada halaman rekomendasi prodi, pengguna tidak hanya dapat melihat daftar program studi yang direkomendasikan, tetapi juga penjelasan singkat mengenai alasan di balik rekomendasi tersebut. Misalnya, system akan menampilkan bahwa siswa yang memiliki minat kuat dalam teknologi dan ilmu computer lebih cocok untuk program studi seperti Teknik informatika atau Sistem informasi, disertai dengan penjelasan mengenai keterkaitan antara minat siswa dan program studi tersebut.

4.7.3 Rekomendasi Prodi Page

Pada halaman rekomendasi page merupakan halaman akhir yang akan menampilkan beberapa informasi. Pertama akan menampilkan nama dari siswa yang di klik dari halaman data siswa sebelumnya, kemudian terdapat juga data minat bakat dari siswa yang bersangkutan, dan rekomendasi prodi dari siswa tersebut .

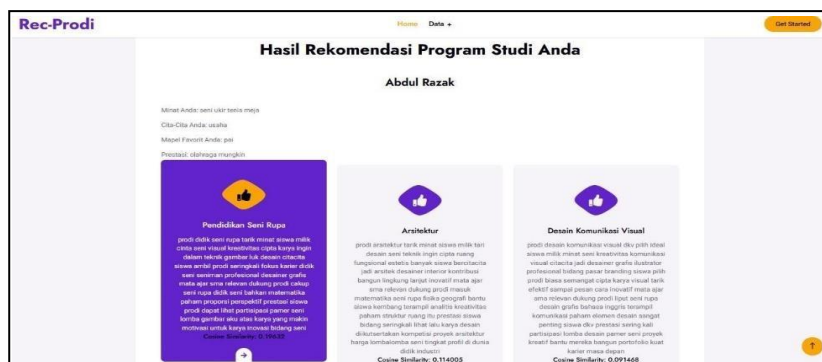
Pada halaman rekomendasi prodi page, pengguna akan melihat informasi terperinci yang merupakan hasil akhir dari proses analisis minat bakat setiap siswa. Halaman ini menampilkan nama siswa yang dipilih dari data siswa page, sehingga pengguna dapat dengan mudah memastikan bahwa rekomendasi yang ditampilkan sesuai dengan profil siswa yang mereka pilih sebelumnya.

Di bawah nama siswa, terdapat bagian yang merangkum data minat bakat siswa tersebut, meliputi minat akademik, cita-cita, mata pelajaran favorit, dan prestasi non-akademik yang diinput pada formulir awal. Ringkasan data ini disajikan dengan jellas agar pengguna, baik siswa, guru, maupun konselor, dapat melihat alasan di balik rekomendasi yang berikan.

Bagian utama dari halaman ini adalah rekomendasi prodi, yang menampilkan daftar program studi yang disarankan untuk siswa berdasarkan analisis profil minat bakatnya, rekomendasi ini disajikan dalam bentuk card yang menarik dan informatif, dengan setiap card mewakili satu program studi yang direkomendasikan. Setiap card menampilkan nama program studi, deskripsi singkat, dan alasan mengapa program tersebut cocok bagi siswa. Misalnya, jika seorang siswa memiliki minat dalam teknologi dan pemrograman, card

rekomendasi mungkin menampilkan program studi seperti Teknik Informatika atau Sistem Informasi, dengan penjelasan singkat mengenai keterkaitan minat siswa dengan kemampuan yang akan dikembangkan dalam program studi tersebut.

Tampilan card dirancang dengan visual yang menarik untuk meningkatkan keterbacaan dan kenyamanan pengguna. Setiap card juga menyertakan ikon atau symbol yang terkait dengan studi yang direkomendasikan, seperti ikon computer untuk bidang teknologi atau ikon buku untuk bidang ilmu Pendidikan. Hal ini bertujuan agar siswa dan pengguna lainnya dapat mengenali dengan cepat bidang dari setiap program studi yang direkomendasikan.



Gambar 4.8 Recommendation Page

Pada halaman rekomendasi ini akan menampilkan beberapa informasi terkait program studi yang direkomendasikan oleh system menggunakan metode content-based filtering. System ini menganalisis data minat dan bakat siswa, kemudian mencocokkannya dengan deskripsi dan atribut dari berbagai program studi untuk menghasilkan rekomendasi yang paling sesuai.

Setiap program studi yang di rekomendasikan ditampilkan dalam bentuk card yang rapi dan informatif, sehingga mudah dibaca dan menarik perhatian. Nama Program Studi ditampilkan di bagian atas card untuk memberikan informasi utama

mengenai pilihan program direkomendasikan. Nama ini dibuat dengan gayayang menonjol, agar pengguna dapat segera mengenali program studi yang mungkin sesuai dengan minat siswa.

Di bawah nama program studi, terdapat Deskripsi Program Studi yang memberikan gambaran singkat mengenai bidang keilmuan yang dipelajari dalam program tersebut, termasuk keterampilan atau pengetahuan yang akan didapatkan siswa jika memilih program ini. Deskripsi ini dirancang untuk memberikan pemahaman awal tentang focus utama dari program studi yang direkomendasikan, sehingga siswa dapat mempertimbangkan apakah bidang ini sesuai dengan minat dan aspirasi mereka. Contohnya, untuk pemrograman, pengembangan perangkat lunak, dan pemahaman teknologi informasi.

Selain itu, card ini juga menampilkan Skor Cosine Similarity yang dihitung oleh system untuk menunjukkan tingkat kecocokan anantara profil minat siswa dengan setiap program studi yang direkomendasikan. Skor Cosine Similarity ini dinyatakan dalam bentuk angka, seperti 0,19 atau 0,11, yang mewakili seberapa tinggi kesesuaian antara minat bakat siswa dan atribut program studi tersebut. Semakin tinggi nilai cosine similarity, semakin tinggi pula kecocokan antara preferensi siswa dengan program studi yang direkomendasikan. Hal ini memberi siswa dan pendamping akademik wawasan lebih kuantitatif mengenai alasan di balik setiap rekomendasi, sehingga mereka dapat mempertimbangkan rekomendasi tersebut secara lebih objektif.

Setiap card juga memiliki ikon visual yang terkait dengan studi yang direkomendasikan, serta tombol “Pelajari Lebih Lanjut” yang dapat diklik untuk

mendapatkan informasi lebih detail tentang program studi tersebut, seperti kurikulum, peluang karier, dan syarat pendaftaran.

Untuk melengkapi informasi ini, halaman rekomendasi prodi juga menyediakan fitur untuk mengurutkan rekomendasi berdasarkan skor cosine similarity sehingga siswa dapat melihat program studi dengan kecocokan tertinggi di bagian atas. Fitur ini dirancang untuk memberikan fleksibilitas dan kemudahan dalam memahami hasil rekomendasi, membantu siswa dalam membuat keputusan yang lebih tepat terkait jalur Pendidikan di masa depan.

4.8 Pembahasan

Sistem rekomendasi berbasis *content-based filtering* merupakan salah satu metode dalam system rekomendasi yang menggunakan pendekatan konten atau karakteristik spesifik pada data untuk menghasilkan rekomendasi yang relevan bagi pengguna. Metode ini berfokus pada analisis data dengan memperhatikan informasi yang diberikan oleh setiap pengguna dan data yang terkait. Dalam penelitian ini, proses rekomendasi dilakukan melalui beberapa tahapan penting, yaitu preprocessing data, pembobotan data menggunakan TF-IDF, dan perhitungan cosine similarity untuk mengukur tingkat kemiripan antara profil minat bakat siswa dan atribut program studi.

Data yang digunakan dalam penelitian ini adalah data minat siswa dan bakat dari siswa-siswi SMA Negeri 1 Sangkapura yang dikumpulkan melalui kuesioner. Secara khusus, penelitian ini memanfaatkan 100 data minat bakat siswa serta 23 data program studi yang diambil dari berbagai referensi di internet. Setiap data minat siswa mencakup elemen-elemen yang relevan, seperti minat siswa, cita-cita, mata

pelajaran favorit, serta prestasi non-akademik yang telah mereka raih. Informasi ini digunakan sebagai dasar untuk mencocokkan profil siswa dengan program studi yang memiliki atribut serupa.

Sistem rekomendasi yang dihasilkan diharapkan mampu membantu pengguna, baik siswa maupun pendamping akademik, dalam memberikan panduan yang tepat mengenai program studi yang sesuai dengan kriteria dan preferensi siswa. Metode ini sangat bermanfaat dalam membantu siswa memahami pilihan program studi yang paling relevan dengan minat dan bakat mereka, sekaligus mendukung visi Pendidikan yang sejalan dengan prinsip ajaran islam, yaitu memberikan nasihat yang baik dan arahan positif kepada sesama. Hal ini sesuai dengan pesan Al-Qur'an, Surah Al-Imran ayat 110, yang menekankan pentingnya memberi petunjuk dan nasihat yang baik kepada umat manusia

كُنْتُمْ خَيْرَ أُمَّةٍ أُخْرِجَتْ لِلنَّاسِ تَأْمُرُونَ بِالْمَعْرُوفِ وَتَنْهَوْنَ عَنِ الْمُنْكَرِ وَتُؤْمِنُونَ بِاللَّهِ ۗ وَلَوْ آمَنَ أَهْلُ الْكِتَابِ لَكَ أَلْكَ أَنْ خَيْرًا لَهُمْ ۗ مِنْهُمْ الْمُؤْمِنُونَ وَأَكْثَرُهُمُ الْفَاسِقُونَ

“Kamu (umat islam) adalah umat terbaik yang dilahirkan untuk (memberi)petunjuk kepada manusia, menyuruh kepada yang ma’ruf, dan mencegah dari yangmunkar, dan beriman kepada Allah SWT” (QS Al-Imran:110).

Ayat ini menegaskan kedudukan umat islam sebagai umat terbaik (khairul ummah ummah) karena tiga sifat utama yang melekat pada mereka menyeru kepada kebaikan (amar ma’ruf), mencegah kemungkaran (nahi munkar), dan beriman kepada allah. Kedudukan ini tidak hanya di anugrahkan begitu saja, melainkan hasil dari kepatuhan umat islam terhadap ajaran agama yang menekankan akhlak mulia, tanggung jawab sosial, dan komitmen terhadap nilai-nilai keimanan.

Menurut Ibnu Katsir pada Tafsir Al-Quran Al Azim Jilid 2 hal 74, keutamaan umat Islam ini didasarkan pada kesempurnaan agama Islam yang diturunkan kepada mereka serta pengalaman ajarannya secara nyata. Tugas menyeru kepada kebaikan dan mencegah kemungkaran menjadi tanggung jawab kolektif untuk menjaga moralitas dan tatanan sosial. Namun, gelar “umat terbaik” hanya berlaku selama umat Islam menjalankan tugas ini dengan konsisten.

Sayyid Qutb dalam *Fi Zilalil Qur'an* menekankan bahwa ayat ini tidak hanya berbicara tentang keutamaan, tetapi juga tanggung jawab besar yang harus dipikul oleh umat Islam. Peran amar ma'ruf nahi munkar menuntun mereka untuk aktif dalam memimpin masyarakat ke arah perbaikan moral dan spiritual, sekaligus menghadapi berbagai tantangan yang mengancam nilai-nilai Islam.

Quraish Shihab dalam tafsir *Al-misbah* menguraikan bahwa keunggulan umat Islam terletak pada kemampuan mereka untuk menjadi teladan yang memberikan manfaat bagi manusia secara universal, bukan hanya kepada sesama Muslim. Hal ini menekankan pentingnya sikap inklusif dan keterlibatan dalam kehidupan sosial, politik, dan ekonomi tanpa mengabaikan nilai-nilai keimanan. Secara umum, ayat ini menegaskan pentingnya iman sebagai fondasi moral serta tugas sosial umat Islam untuk berkontribusi aktif dalam membangun masyarakat yang adil dan harmonis. Ayat ini juga mengingatkan umat Islam agar tidak bersikap pasif atau eksklusif, tetapi menjadi agen perubahan yang membawa kebaikan universal.

Dalam konteks pencarian program studi, sistem rekomendasi berperan sebagai alat yang efektif untuk memberikan panduan yang tepat bagi seseorang

dalam menemukan program studi yang selaras dengan minat dan bakat mereka. sistem ini tidak hanya membuat siswa dalam memilih jalur Pendidikan yang sesuai, tetapi juga berfungsi sebagai sarana untuk mengenali potensi diri yang mungkin belum sepenuhnya disadari. Melalui analisis data minat dan bakat, sistem rekomendasi dapat menyajikan pilihan program studi yang paling mendekati profil pengguna, sehingga memudahkan siswa dalam mengambil keputusan yang berdasarkan preferensi dan potensi pribadi mereka.

Sebagaimana yang di ajarkan dalam agama islam, memberikan nasihat dan petunjuk yang baik kepada sesama adalah perbuatan mulia)ma'ruf(yang di anjurkan unntuk menciptakan mafaat dan kebijakan bagi umat. Prinsip ini sejalan dengan pentingnya membantu siswa menemukan bidang studi yang cocok, karena pilihan Pendidikan adalah langkah penting dalam membentuk masa depan dan potensi diri seseorang. Islam mendorong umatnya untuk saling memberi petunjuk dan membantu dalam kebaikan, sebagaimana firamat Allah dalam QS. Al-Imran:100 yang menyatakan bahwa umat islam adalah sebaik-baiknya umat karena menyeru kepada yang ma'ruf dan mencegah dari yang munkar. Dalam hal ini, sistem rekomendasi program studi dapat dianggap sebagai wujud nyata dari prinsip ma'ruf ini, di mana etknologi digunakan untuk memberiuikan manfaat langsung dalam kehidupan sehari-hari.

Dengan memanfaatkan teknologi sistem rekomendasi, seseorang dapat menemukan program studi yang paling sesuai dengan kriteria, bakat, dan minat yang mereka miliki. Hal ini memungkinkan terciptanya keselarasan antara minat yang mereka miliki. Hal ini memungkinkan terciptanya keselarasan antara minat siswa yang mereka miliki. Ini memungkinkan terciptanya keselarasan antara minat siswa

dengan program studi yang akan mereka pilih, sehingga diharapkan mereka dapat belajar dengan lebih antusias dan berprestasi. Keserasalan antara minat dan pilihan studi juga dapat meminimalisir kemungkinan siswa merasa terpaksa atau salah jurusan di kemudian hari, karena mereka telah dipandu oleh sistem rekomendasi yang mempertimbangkan preferensi pribadi mereka.

Selain itu, pengguna sistem rekomendasi dalam konteks Pendidikan juga memberikan keuntungan jangka Panjang bagi perkembangan Pendidikan itu sendiri. Dengan membantu siswa menemukan jalur yang sesuai dengan minat mereka, sistem ini dapat berkontribusi pada pembentukan generasi yang lebih terarah dan bersemangat dalam mengajar ilmu yang benar-benar mereka sukai. Dengan demikian, teknologi ini tidak hanya bermanfaat secara individu, tetapi juga membawa dampak positif bagi komunitas Pendidikan secara keseluruhan, menciptakan lingkungan belajar yang lebih inklusif dan responsive terhadap kebutuhan setiap siswa.

Pada akhirnya, sistem rkeomendasi dibangun atas dasar content-based filtering ini mampu membantu siswa Dallam mencapai potensi terbaik mereka. Dukungan teknologi metode seperti ini menjadi sarana yang mendekatkan antara minat, bakat, dan pilihah Pendidikan siswa, sehingga dapat mendukung tercapainya tujuan hidup yang lebih baik sesuai dengan prinsip-prinsip petunjuk dan nasihat baik dalam islam.

Sistem rekomendasi yang berhasil dikembangkan dalam penelitian ini memanfaatkan sejumlah kriteria atau preferensi minat dan bakat siswa sebagai dasar untuk menentukan hasil rekomendasi yang relevan. Beberapa kriteria utama

yang digunakan meliputi minat siswa, cita-cita atau tujuan karier, mata pelajaran favorit, dan prestasi non-akademik yang pernah diraih siswa. Setiap kriteria ini dipilih dengan cermat karena dianggap dapat mencerminkan minat dan potensi siswa, yang menjadi landasan penting dalam proses pemilihan program studi. Minat siswa, misalnya, memberikan gambaran umum tentang bidang atau disiplin ilmu yang menarik perhatian mereka, seperti teknologi, seni, ilmu sosial, atau sains. Cita-cita atau tujuan karier mencerminkan aspirasi jangka panjang siswa, yang dapat diarahkan pada profesi tertentu, seperti dokter, insinyur, guru, atau pengusaha. Informasi ini membantu sistem untuk mengarahkan rekomendasi program studi yang relevan dengan tujuan karier siswa, sehingga pilihan program studi yang direkomendasikan tidak hanya sesuai dengan minat akademik, tetapi juga mendukung pencapaian karier di masa depan.

Selanjutnya, mata pelajaran favorit menunjukkan bidang akademik yang secara alami lebih menarik bagi siswa dan di mana mereka mungkin memiliki keterampilan atau minat khusus, seperti matematika, bahasa, atau ilmu alam. Data ini menjadi acuan dalam penyaringan program studi yang sesuai, sehingga siswa mendapatkan rekomendasi jurusan yang tidak hanya relevan dengan cita-cita mereka tetapi juga berpotensi mendukung kekuatan akademik mereka. Prestasi non-akademik, seperti kejuaraan olahraga, penghargaan dalam kompetisi seni, atau keterlibatan dalam organisasi, juga diperhitungkan karena dianggap mewakili tambahan dan keterampilan siswa di luar bidang akademik. Hal ini membantu sistem memberikan rekomendasi yang lebih komprehensif dan sesuai dengan profil siswa secara keseluruhan. Setiap kriteria atau preferensi ini kemudian diolah

menggunakan perhitungan cosine similarity. Perhitungan cosine similarity ini berfungsi untuk mengukur tingkat kemiripan antara profil minat dan bakat siswa dengan atribut dari setiap program studi yang tersedia dalam sistem. Semakin tinggi nilai cosine similarity yang dihasilkan, semakin cocok atau relevan program studi tersebut dengan profil siswa. Misalnya, jika seorang siswa menunjukkan minat yang kuat dalam seni dan memiliki cita-cita sebagai desainer, maka program studi seperti Desain Komunikasi Visual atau Seni Rupa akan memiliki skor cosine similarity yang lebih tinggi dibandingkan dengan program studi yang tidak terkait.

Hasil perhitungan cosine similarity ini menjadi dasar utama dalam sistem rekomendasi, di mana program studi dengan skor kemiripan tertinggi akan direkomendasikan terlebih dahulu kepada siswa. Dengan demikian, sistem ini tidak hanya menghasilkan rekomendasi yang sesuai dengan minat siswa, tetapi juga mempertimbangkan beragam faktor personal yang unik bagi setiap siswa. Melalui pendekatan yang mempertimbangkan preferensi dan minat secara spesifik, sistem ini diharapkan dapat membantu siswa membuat pilihan program studi yang lebih personal dan sesuai dengan potensi mereka, mendukung kesuksesan akademik dan profesional di masa depan.

Pada Gambar 4.8, diperoleh data siswa bernama Abdul Rozak, yang memiliki sejumlah minat dan bakat yang beragam. Abdul menunjukkan minat khusus pada seni ukir dan tenis meja, dengan cita-cita menjadi seorang pengusaha. Mata pelajaran favoritnya adalah Pendidikan Agama Islam (PAI), dan ia juga memiliki prestasi dalam bidang olahraga. Berdasarkan data minat dan bakat ini, sistem rekomendasi menghasilkan tiga program studi dengan skor tertinggi yang

paling sesuai dengan profil Abdul, yaitu:

Program studi Pendidikan seni rupa dengan skor 0.196, program studi Arsitektur dengan skor 0.114, dan program studi Desain Komunikasi Visual dengan skor 0.091. Di antara ketiga rekomendasi ini, program studi Pendidikan Seni Rupa memiliki skor tertinggi, mendekati nilai 0.2, yang menunjukkan kecocokan paling tinggi dengan profil minat dan bakat Abdul dibandingkan dua program studi lainnya. Kesesuaian yang tinggi ini terutama didasarkan pada adanya kata kunci “seni” yang mendominasi profil minat Abdul, sehingga program studi Pendidikan Seni Rupa menonjol dalam hasil rekomendasi. Selain itu, Pendidikan Seni Rupa mencakup elemen-elemen yang sejalan dengan minat Abdul dalam seni ukir, serta keterkaitan dengan cita-cita berwirausaha, yang dapat diperkuat melalui keterampilan seni visual.

Skor rekomendasi yang dihasilkan dihitung berdasarkan tingkat kemiripan profil Abdul dengan atribut yang dimiliki oleh masing-masing program studi, menggunakan cosine similarity untuk mengukur kesesuaian. Meski skor 0.196 belum mendekati angka 1, program studi Pendidikan Seni Rupa memiliki keterkaitan yang cukup jelas dengan minat dan potensi yang dimiliki Abdul, sehingga menjadi pilihan utama dalam daftar rekomendasi. Adapun skor untuk program studi Arsitektur dan Desain Komunikasi Visual juga memiliki kemiripan yang signifikan, walaupun tidak sebesar Pendidikan Seni Rupa. Program Arsitektur mendapatkan skor 0.114 karena memiliki keterkaitan dengan elemen desain dan visualisasi, yang berkaitan dengan minat Abdul dalam seni, walaupun lebih terfokus pada aspek teknis dan struktural. Sementara itu, program Desain Komunikasi

Visual mendapatkan skor 0.091, yang masih cukup relevan karena juga mengandung unsur seni dan desain. Namun, karena minat Abdul lebih pada seni ukir dan aspek visual tradisional, Pendidikan Seni Rupa dianggap sebagai rekomendasi yang paling cocok.

Analisis hasil ini menunjukkan bahwa nilai cosine similarity yang diperoleh belum mendekati 1, yang berarti terdapat ruang untuk meningkatkan kesesuaian data input minat dan bakat siswa dengan deskripsi program studi yang tersedia. Data minat dan bakat yang lebih spesifik atau lebih lengkap mengenai keterampilan akademik siswa juga akan membantu meningkatkan akurasi dan kualitas rekomendasi, memungkinkan sistem untuk memberikan hasil yang lebih mendekati kecocokan optimal dengan minat siswa seperti Abdul Rozak.

Di samping itu, hasil yang direkomendasikan pada Gambar 4.8 menunjukkan beberapa kekurangan yang signifikan. Skor cosine similarity yang diperoleh tergolong rendah dan masih jauh dari angka 1, yang menandakan adanya ketidaksesuaian dalam proses pencocokan data. Rendahnya nilai ini bisa disebabkan oleh beberapa faktor, salah satunya adalah relevansi data input mengenai minat dan bakat yang digunakan dalam sistem. Ketidakcocokan ini dapat mengindikasikan bahwa data yang dimasukkan tidak sepenuhnya mencerminkan potensi dan kemampuan akademik siswa, sehingga rekomendasi yang dihasilkan juga cenderung memiliki nilai yang cukup kecil. Oleh karena itu, penting untuk melakukan evaluasi dan perbaikan pada metode pengumpulan serta analisis data input agar rekomendasi yang diberikan dapat lebih akurat dan sesuai dengan kemampuan serta minat siswa.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini berhasil membangun sistem rekomendasi program studi yang ditujukan bagi siswa Kelas 12 SMA Negeri 1 Sangkapura dengan menerapkan metode *content-based filtering*.

Sistem rekomendasi berbasis *Content-Based Filtering* telah terbukti efektif dalam membantu siswa memilih program studi yang sesuai dengan minat dan bakat mereka. Dengan memanfaatkan data profil siswa, seperti minat, cita-cita, mata pelajaran favorit, dan prestasi non-akademik, sistem ini mampu memberikan rekomendasi program studi yang relevan. Proses ini melibatkan tahapan utama, yaitu *preprocessing* data, pembobotan data menggunakan metode TF-IDF, dan perhitungan *cosine similarity* untuk mengukur tingkat kecocokan antara profil siswa dan atribut program studi.

Penelitian ini menyoroti pentingnya keselarasan antara pilihan pendidikan siswa dengan potensi pribadi mereka, sehingga diharapkan dapat meningkatkan motivasi belajar dan mengurangi risiko salah jurusan. Contoh kasus yang melibatkan siswa bernama Abdul Rozak menunjukkan bahwa program studi yang paling sesuai, seperti Pendidikan Seni Rupa, dipilih berdasarkan nilai *cosine similarity* tertinggi. Meski demikian, nilai kecocokan yang dihasilkan masih dapat ditingkatkan melalui penyempurnaan data input dan algoritma.

Sistem ini tidak hanya memberikan manfaat individu, tetapi juga berdampak positif pada dunia pendidikan secara umum dengan menciptakan lingkungan belajar yang inklusif dan responsif terhadap kebutuhan siswa. Sebagai implementasi nyata dari prinsip amar ma'ruf nahi munkar dalam Islam, sistem rekomendasi ini mendorong keterlibatan aktif dalam memberikan nasihat dan petunjuk yang baik. Prinsip ini sejalan dengan ajaran Islam yang menekankan pentingnya membantu sesama dalam mencapai potensi terbaik mereka, sebagaimana ditegaskan dalam QS. Ali Imran: 110.

Dengan adanya pengembangan lebih lanjut, seperti integrasi metode *Collaborative Filtering* atau *machine learning*, kualitas dan akurasi sistem rekomendasi ini dapat ditingkatkan. Sistem ini diharapkan mampu menjadi alat strategis yang mendukung siswa dalam mencapai kesuksesan akademik dan profesional di masa depan, sekaligus membangun generasi yang lebih berorientasi pada potensi dan minat yang autentik.

Proses pengembangan sistem rekomendasi berbasis web melalui beberapa tahap kunci, dimulai dari preprocessing data, transformasi ke matriks TF-IDF, hingga perhitungan kemiripan menggunakan algoritma Cosine Similarity.

1. **Transformasi TF-IDF:** Data teks diubah menjadi representasi numerik dengan `TfidfVectorizer`, yang memberi bobot lebih besar pada kata-kata yang jarang muncul di koleksi dokumen tetapi relevan dalam dokumen tertentu. Hasilnya adalah matriks TF-IDF, yang menjadi dasar untuk analisis lebih lanjut.
2. **Perhitungan Kemiripan (Cosine Similarity):** Setelah matriks TF-IDF

terbentuk, kemiripan antar dokumen dihitung dengan algoritma Cosine Similarity. Algoritma ini membandingkan vektor dokumen untuk mengevaluasi relevansi, di mana nilai mendekati 1 menunjukkan kemiripan tinggi. Dalam implementasi, proses ini dilakukan dengan menghitung dot product dan panjang vektor dokumen.

3. **Hasil Analisis:** Dari hasil analisis, ditemukan bahwa sebagian besar nilai bobot antar dokumen adalah nol, mengindikasikan rendahnya kesamaan antara query dengan dokumen. Namun, dokumen tertentu dengan nilai kemiripan lebih tinggi dianggap lebih relevan.
4. **Implementasi Sistem Rekomendasi:** Sistem rekomendasi dikembangkan menggunakan framework Flask, yang mengintegrasikan Python untuk pengolahan data dengan HTML, CSS, dan JavaScript untuk tampilan antarmuka. Aplikasi ini menyediakan rekomendasi program studi yang relevan berdasarkan analisis data, dengan visualisasi yang intuitif dan responsif untuk berbagai perangkat.

Keseluruhan proses ini menunjukkan pentingnya metode TF-IDF dan Cosine Similarity dalam membangun sistem rekomendasi berbasis konten yang efektif dan user-friendly.

5.2 Saran

Berikut merupakan beberapa saran untuk penelitian serupa di masa yang akan datang. Saran-saran ini didasarkan dari hasil implementasi data, serta metode yang digunakan dalam penelitian ini. Tujuan dari saran-saran ini adalah untuk memperbaiki performa dan akurasi sistem rekomendasi serta memastikan bahwa

hasil rekomendasi lebih relevan dan tepat sasaran. Rekomendasi tersebut antara lain:

1. Pengisian Data Minat dan Bakat yang Lebih Mendalam

Pengisian data minat dan bakat oleh siswa dapat diperbaiki dengan mencakup deskripsi yang lebih mendalam mengenai minat dan bakat akademik maupun non-akademik. Hal ini penting agar sistem dapat memahami minat siswa dengan lebih baik, meningkatkan relevansi antara data minat siswa dan deskripsi program studi, sehingga rekomendasi yang diberikan lebih terarah. Penggunaan deskripsi terstruktur pada form, seperti pengelompokan bidang akademik (sains, teknologi, seni, dll.), dapat mempermudah sistem dalam mengidentifikasi kecocokan minat siswa dengan program studi.

2. Penyediaan Panduan Pengisian Formulir Google Forms

Sebaiknya disediakan panduan pengisian formulir yang jelas dan mendetail untuk mengarahkan siswa dalam memberikan jawaban yang lebih akurat dan sesuai dengan kebutuhan sistem rekomendasi. Panduan ini dapat mencakup contoh jawaban atau kata kunci terkait minat dan bakat yang sesuai dengan berbagai program studi, sehingga siswa memiliki gambaran yang lebih jelas tentang informasi yang diharapkan oleh sistem. Dengan panduan ini, siswa dapat lebih mudah menyampaikan minat mereka dengan cara yang dapat diproses secara efektif oleh algoritma rekomendasi.

3. Eksplorasi Lebih Lanjut tentang Metode dalam Sistem Rekomendasi

Untuk meningkatkan akurasi dan performa sistem, disarankan untuk melakukan eksplorasi lebih lanjut mengenai metode sistem rekomendasi yang lain. Contohnya,

metode hybrid filtering, yang mengombinasikan content-based filtering dengan collaborative filtering, dapat digunakan untuk memperkuat hasil rekomendasi. Dengan hybrid filtering, sistem dapat memanfaatkan data historis siswa lain yang memiliki profil serupa untuk meningkatkan akurasi rekomendasi, terutama jika terdapat keterbatasan pada data minat dan bakat.

4. Peningkatan Jumlah Data Program Studi

Disarankan untuk menambah jumlah dan variasi data program studi agar lebih mencakup berbagai minat siswa. Data yang lebih kaya dan beragam akan memungkinkan sistem rekomendasi memberikan pilihan program studi yang lebih spesifik dan sesuai dengan minat unik siswa. Hal ini dapat dilakukan dengan mengintegrasikan data dari beberapa universitas atau platform yang memiliki informasi program studi yang lengkap dan terperinci.

5. Peningkatan Kualitas Data dengan Validasi dan Klasifikasi Minat

Untuk memperoleh data input yang lebih akurat, disarankan agar data minat siswa divalidasi dan dikelompokkan ke dalam kategori minat yang jelas sebelum diproses. Pengelompokan ini dapat dilakukan dengan bantuan alat validasi berbasis teks atau klasifikasi manual, yang bertujuan untuk menyesuaikan jawaban siswa dengan kategori minat yang tersedia dalam sistem. Langkah ini akan membantu meningkatkan kesesuaian antara profil siswa dan rekomendasi yang dihasilkan, serta mengurangi kemungkinan hasil rekomendasi yang kurang relevan.

6. Pengembangan Visualisasi Hasil Rekomendasi untuk Pengguna

Sebagai bentuk peningkatan antarmuka, disarankan agar sistem rekomendasi menyajikan visualisasi hasil rekomendasi yang lebih interaktif. Misalnya, dengan

menampilkan grafik yang menunjukkan tingkat kecocokan setiap program studi atau tabel perbandingan program studi berdasarkan beberapa parameter utama. Visualisasi ini akan membantu siswa dan pendamping akademik memahami pilihan yang tersedia dengan lebih mudah, serta menambah daya tarik dan kejelasan informasi rekomendasi.

7. Pengembangan sistem rekomendasi program studi

Penelitian ini mencatat beberapa keterbatasan yang perlu diatasi untuk pengembangan di masa depan. Keterbatasan utama meliputi kurangnya keragaman data referensi program studi dan kelemahan pada formulir pengisian data minat dan bakat, yang mengurangi akurasi rekomendasi sistem. Untuk mengatasinya, diperlukan langkah seperti memperluas data referensi, menyusun panduan pengisian formulir yang terstruktur, serta menambahkan fitur umpan balik dari siswa. Pengembangan ini diharapkan meningkatkan performa dan akurasi sistem berbasis content-based filtering, sehingga menjadi alat bantu yang lebih efektif dan komprehensif dalam mendukung keputusan pendidikan siswa serta memberikan dampak positif bagi sektor pendidikan secara lebih luas.

Dengan menerapkan saran-saran di atas, diharapkan bahwa sistem rekomendasi di masa mendatang dapat memberikan hasil rekomendasi yang lebih akurat, relevan, dan bermanfaat bagi siswa. Hal ini sekaligus mendukung tujuan utama sistem rekomendasi, yaitu memberikan panduan yang tepat bagi siswa dalam memilih program studi yang sesuai dengan potensi dan aspirasi akademik mereka

DAFTAR PUSTAKA

- Adomavicius, G., & Tuzhilin, A. (2018). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734–749. <https://doi.org/10.1109/TKDE.2005.99>
- Amri, S., Jauhari, A., & Elisah, T. (2016). *Implementasi Pendidikan Karakter dalam Pembelajaran*. Prestasi Pustaka.
- Cui, Z., Xu, X., Xue, F., Cai, X., Cao, Y., Zhang, W., & Chen, J. (2020).
- Februariyanti, H., Laksono, A. D., Wibowo, J. S., & Utomo, M. S. (2021). Implementasi Metode Collaborative Filtering Untuk Sistem Rekomendasi Penjualan Pada Toko Mebel. *Jurnal Khatulistiwa Informatika*, 9(1), 43–50. <https://doi.org/https://doi.org/10.31294/jki.v9i1.9859>
- Huda, A. A., Fajarudin, R., & Hadinegoro, A. (2022). Sistem Rekomendasi Content-based Filtering Menggunakan TF-IDF Vector Similarity Untuk Rekomendasi Artikel Berita. *Building of Informatics, Technology and Science (BITS)*, 4(3). <https://doi.org/10.47065/bits.v4i3.2511>
- Larasati, F. B. A., & Februariyanti, H. (2021). Sistem Rekomendasi Product EminaCosmetics Dengan Menggunakan Metode Content-Based Filtering. *Jurnal Manajemen Informatika Dan Sistem Informasi*, 4(1), 45. <https://doi.org/10.36595/misi.v4i1.250>
- Nastiti, P. (2019). Penerapan Metode Content Based Filtering Dalam Implementasi Sistem Rekomendasi Tanaman Pangan. *Teknika*, 8(1), 1–10. <https://doi.org/10.34148/teknika.v8i1.139>
- Personalized Recommendation System Based on Collaborative Filtering for IoT Scenarios. *IEEE Transactions on Services Computing*, 13(4), 685–695. <https://doi.org/10.1109/TSC.2020.2964552>
- Prayoga, F. A., & Kusnawi, K. (2022). Smartphone Recommendation System Using Model-Based Collaborative Filtering Method. *Jurnal Teknik Informatika (Jutif)*, 3(6), 1613–1622. <https://doi.org/10.20884/1.jutif.2022.3.6.413>
- Putri, I. H., Muslimah, S. K. H., Nurakhmadyavi, & Wahyudi, E. E. (2022). Literature Review: Sistem Rekomendasi untuk Buku dan Film. *Prosiding Seminar Nasional Teknik Elektro, Sistem Informasi, Dan Teknik Informatik (SNESTIK)*, 1(1), 75–81. <https://doi.org/https://doi.org/10.31284/p.snestic.2022.2635>

- Rahmadani, A. T., & Arif, M. (2017). Pengaruh Faktor Internal Dan faktor Exsternal Terhadap Minat Siswa Memilih Jurusan IPS Di Kelas Xi Sma Negeri 1 Kotapinang Kabupaten Labuhanbatu Selatan. *JURNAL GEOGRAFI*, 9(1), 78. <https://doi.org/10.24114/jg.v9i1.6040>
- Ricardo, R., & Meilani, R. I. (2017). Impak Minat dan Motivasi Belajar Terhadap Hasil Belajar Siswa. *Jurnal Pendidikan Manajemen Perkantoran*, 2(2), 79. <https://doi.org/10.17509/jpm.v2i2.8108>
- Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (2020). *Recommender Systems Handbook*. Springer US. <https://doi.org/10.1007/978-0-387-85820-3>
- Rifaldy, R., & Setiawan, E. B. (2022). Recommender System Movie Netflix using Collaborative Filtering with Weighted Slope One Algorithm in Twitter. *Building of Informatics, Technology and Science (BITS)*, 4(2), 500–506. <https://doi.org/10.47065/bits.v4i2.1959>
- Rufaidah, A. (2015). Pengaruh Intlegensi Dan Minat Siswa Terhadap Putusan Pemilihan Jurusan. *Faktor: Jurnal Ilmiah Kependidikan*, 2(2), 139–146.
- Zhang, S., Yao, L., Sun, A., & Tay, Y. (2020). Deep Learning Based Recommender System. *ACM Computing Surveys*, 52(1), 1–38. <https://doi.org/10.1145/3285029>
- Zhang, W., Zhang, X., Wang, H., & Chen, D. (2019). A deep variational matrix factorization method for recommendation on large scale sparse dataset. *Neurocomputing*, 334, 206–218. <https://doi.org/10.1016/j.neucom.2019.01.028>