

# PENERAPAN METODE CONTENT-BASED FILTERING PADA SISTEM REKOMENDASI KURSUS CODING

Fauzi Nur Azis<sup>1\*</sup>, Bondan Wahyu Pamekas<sup>2</sup>, Hanifah Permatasari<sup>3</sup>

<sup>1</sup>Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta  
<sup>1\*</sup>220103058@mhs.udb.ac.id

<sup>2</sup>Teknologi Rekayasa Perangkat Lunak/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta  
<sup>2</sup>bondan\_wahyupamekas@udb.ac.id

<sup>3</sup>Sistem Informasi/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta  
<sup>3</sup>hanifah\_permatasari@udb.ac.id

**Abstrak**— Perkembangan teknologi digital telah memperluas akses pembelajaran coding melalui berbagai platform daring. Namun, banyaknya pilihan kursus dengan topik dan tingkat pembelajaran yang beragam sering membuat pengguna kesulitan menentukan kursus yang sesuai dengan kebutuhan belajarnya. Tujuan utama penelitian ini adalah menerapkan metode *Content-Based Filtering* pada sistem rekomendasi kursus coding berbasis web. Sistem yang dikembangkan diharapkan mampu memberikan saran kursus yang sesuai dengan kebutuhan belajar pengguna berdasarkan karakteristik konten yang tersedia. Mekanisme rekomendasi pada penelitian ini memanfaatkan pembobotan TF-IDF untuk merepresentasikan fitur teks setiap kursus, kemudian menghitung tingkat kemiripannya menggunakan *Cosine Similarity*. Hasil perhitungan tersebut digunakan sebagai dasar dalam menentukan urutan rekomendasi kursus yang memiliki karakteristik konten paling relevan. Data penelitian ini berupa dataset publik kursus online pada domain *web development* yang telah diseleksi sesuai kebutuhan penelitian. Hasil penelitian menunjukkan bahwa TF-IDF mampu membentuk fitur dominan pada data kursus, sedangkan *Cosine Similarity* digunakan untuk mengurutkan kursus berdasarkan nilai kemiripan konten. Evaluasi sistem memperoleh skor akurasi sebesar 88,00%, presisi 86,67%, dan *recall* 78,57%. sehingga metode *Content-Based Filtering* dapat digunakan untuk membantu pengguna dalam memilih kursus coding yang sesuai.

**Kata kunci**— *content-based filtering*, kursus coding, *web development*, TF-IDF, *cosine similarity*, sistem rekomendasi

**Abstract**— The advancement of digital technology has significantly expanded access to programming education via various online platforms. Nevertheless, the abundance of course options with diverse topics and learning levels frequently creates difficulties for users in identifying programs that align with their educational needs. The primary objective of this study is to deploy the Content-Based Filtering method within a web-based coding course recommendation system. The engineered system is projected to supply appropriate course suggestions tailored to the user's learning requirements based on available content characteristics. The recommendation mechanism in this research utilizes TF-IDF weighting to represent the text features of each course, subsequently calculating proximity metrics through Cosine Similarity. The outputs of these calculations serve as the foundation for establishing the recommendation sequence for courses possessing the most relevant content attributes. The research data comprises a public online course dataset within the web development domain, which has been curated to match the study's requirements. The experimental outcomes indicate that TF-IDF successfully extracts dominant features from the course data, while Cosine Similarity is employed to rank the courses according to content similarity scores. System evaluation yields an accuracy score of 88.00%, precision of 86.67%, and recall of 78.57%. Consequently, the Content-Based Filtering framework proves viable in assisting users to select well-suited programming courses.

**Keywords**— *content-based filtering*, coding course, *web development*, TF-IDF, *cosine similarity*, recommendation system.

## I. PENDAHULUAN

Perkembangan teknologi digital mendorong peningkatan kebutuhan akan keterampilan pemrograman di berbagai bidang, sehingga minat mempelajari coding melalui platform daring terus mengalami pertumbuhan. Berbagai sumber pembelajaran, seperti video tutorial, artikel, dan kursus online, tersedia secara luas dan dapat diakses oleh mahasiswa maupun pembelajar mandiri. Ketersediaan kursus yang beragam tersebut di satu sisi memberikan kemudahan akses, namun di sisi lain

menimbulkan kesulitan dalam menentukan kursus yang sesuai dengan kebutuhan serta tingkat kemampuan pengguna. Kesulitan dalam proses pemilihan kursus ini juga ditemukan dalam penelitian sebelumnya, di mana banyaknya pilihan kursus dengan topik dan tingkat kesulitan yang beragam menyebabkan pengguna mengalami kebingungan dalam menentukan kursus yang sesuai dengan kebutuhannya[1].

Metode *Content-Based Filtering* memanfaatkan informasi yang terdapat pada

konten sebagai dasar untuk menentukan item yang memiliki tingkat keterkaitan tinggi dengan preferensi individu [2]. Pada penelitian sistem rekomendasi penelusuran buku, penggunaan metode tersebut menunjukkan bahwa teknik pembobotan teks mampu meningkatkan kualitas rekomendasi melalui proses pencocokan karakteristik konten yang sesuai dengan kebutuhan pengguna [3]. Penelitian lain yang mengimplementasikan metode *Content-Based Filtering* pada aplikasi rekomendasi destinasi wisata berbasis web menunjukkan bahwa proses pengukuran kesamaan karakteristik konten dapat menghasilkan saran yang relevan dengan kebutuhan pengguna [4]. Pendekatan tersebut memiliki kelebihan karena mampu menyesuaikan rekomendasi berdasarkan karakteristik item yang dipilih tanpa memerlukan riwayat pengguna lain. Namun demikian, keberagaman rekomendasi yang dihasilkan masih bergantung pada atribut konten yang tersedia pada sistem [5].

Pada penelitian ini, mekanisme rekomendasi memanfaatkan metode TF-IDF untuk merepresentasikan karakteristik teks dari setiap kursus. Selanjutnya, tingkat kedekatan antar kursus dihitung menggunakan metode *Cosine Similarity* berdasarkan bobot yang telah diperoleh. Melalui tahapan tersebut, sistem dapat mengidentifikasi kursus yang memiliki karakteristik konten serupa sehingga rekomendasi yang dihasilkan menjadi lebih relevan dengan kebutuhan pembelajaran.

Berdasarkan permasalahan yang telah dipaparkan, penelitian ini mengkaji penerapan metode *Content-Based Filtering* untuk membangun sistem rekomendasi kursus coding berbasis web. Sumber data penelitian berasal dari dataset publik kursus daring pada bidang *web development* yang telah dipilih dan disesuaikan dengan kebutuhan penelitian. Proses rekomendasi memanfaatkan atribut judul kursus sebagai dasar dalam menentukan tingkat kemiripan konten, sedangkan informasi mengenai level pembelajaran dan URL disajikan sebagai informasi pendukung pada hasil rekomendasi. Melalui pengembangan sistem tersebut, pengguna diharapkan dapat

menemukan kursus coding yang lebih relevan dengan kebutuhan dan tujuan pembelajarannya.

## II. METODOLOGI PENELITIAN

### A. Jenis dan Sumber Data

Penelitian ini memanfaatkan data sekunder sebagai sumber utama penelitian. Data sekunder merupakan informasi yang telah tersedia sebelumnya sehingga dapat digunakan kembali untuk mendukung kegiatan penelitian. Sumber data yang digunakan berupa dataset publik kursus daring pada bidang web development. Dataset tersebut memiliki tiga atribut, yaitu course title, level, dan url.

Atribut course title dimanfaatkan sebagai komponen utama pada mekanisme rekomendasi karena memuat informasi mengenai topik atau materi yang dipelajari dalam kursus. Sementara itu, atribut level digunakan sebagai informasi tambahan untuk menunjukkan tingkat kesulitan pembelajaran, sedangkan atribut url berfungsi sebagai tautan menuju halaman kursus yang direkomendasikan. Pada tahap pengolahan menggunakan metode TF-IDF dan Cosine Similarity, atribut yang dianalisis hanya course title, sedangkan level dan url tidak dilibatkan dalam proses pengukuran kemiripan.

Dalam penelitian ini, proses pengumpulan data dilakukan melalui dokumentasi dan studi literatur. Penjelasan terkait metode yang digunakan dapat diuraikan sebagai berikut:

#### a. Dokumentasi

Dokumentasi dilakukan dengan menggunakan dataset publik kursus online yang telah tersedia. Data kemudian diseleksi sesuai kebutuhan penelitian, yaitu data yang memiliki atribut course title, level, dan url.

#### b. Studi literatur

Sebagai dasar penyusunan penelitian, peneliti melakukan telaah pustaka terhadap berbagai referensi yang berhubungan dengan topik yang dibahas. Sumber yang digunakan mencakup buku, jurnal ilmiah, dan literatur pendukung lainnya yang relevan. Kajian pustaka tersebut berfungsi untuk memperkuat pemahaman teoritis, meninjau penelitian yang telah dilakukan sebelumnya, serta memberikan gambaran mengenai perkembangan penelitian di bidang sistem rekomendasi. Selain itu, hasil

telaah digunakan untuk menunjukkan kontribusi penelitian terhadap pengembangan sistem rekomendasi kursus coding berbasis web.

Penelitian yang dilakukan oleh Ziliwu mengembangkan platform pemberi saran produk yang memanfaatkan pendekatan *Collaborative Filtering* serta teknologi *Machine Learning*. Hasil pengujian memperlihatkan bahwa rekomendasi produk dapat dihasilkan melalui analisis pola aktivitas pengguna yang memiliki karakteristik serupa serta dinilai efektif dalam mendukung peningkatan performa penjualan secara online. Hasil penelitian mengindikasikan bahwa pendekatan rekomendasi layak diterapkan dalam konteks produk dengan memanfaatkan interaksi dan perilaku pengguna sebagai dasar rekomendasi [6].

Penelitian lain oleh Pratama mengembangkan sistem rekomendasi buku menggunakan pendekatan *User-Based Collaborative Filtering*. Hasil penelitian memperlihatkan bahwa rekomendasi buku dapat dihasilkan dengan memanfaatkan kesamaan pola penilaian antar pengguna. Meskipun demikian, penelitian ini juga mengidentifikasi kendala seperti *sparsity* dan *cold-start* yang masih perlu diperhatikan pada pengembangan selanjutnya. Temuan tersebut memperlihatkan bahwa pendekatan *Collaborative Filtering* cukup efektif dalam menghasilkan rekomendasi, tetapi masih memiliki keterbatasan ketika data interaksi pengguna belum memadai [7].

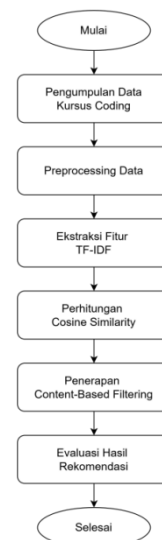
Selanjutnya, Akhdan dan Fitriani melalui penelitian mengenai platform pencarian mentor belajar yang menerapkan pendekatan *Content-Based Filtering* dengan dukungan teknik TF-IDF dan *Cosine Similarity* menunjukkan bahwa sistem mampu merekomendasikan mentor berdasarkan kata kunci keahlian. Hasil penelitian tersebut memperlihatkan bahwa pemilihan mentor dapat dilakukan melalui pencocokan kata kunci kompetensi yang dimiliki. Temuan penelitian juga menunjukkan bahwa kombinasi *Content-Based Filtering*, TF-IDF, dan *Cosine Similarity* efektif dalam mendukung proses pencarian secara lebih tepat dan personal. Penelitian ini memiliki kedekatan dengan penelitian yang dilakukan karena sama-

sama memanfaatkan karakteristik konten sebagai dasar rekomendasi [8].

Berdasarkan beberapa penelitian terdahulu tersebut, dapat diketahui bahwa sistem rekomendasi telah diterapkan pada berbagai domain dengan pendekatan yang berbeda. Penelitian ini berfokus pada penerapan *Content-Based Filtering* yang dipadukan dengan teknik TF-IDF dan *Cosine Similarity* dalam pengembangan sistem rekomendasi kursus coding berbasis web. Perbedaan penelitian ini terletak pada objek rekomendasi dan atribut konten yang digunakan sebagai dasar pencocokan.

### B. Tahapan Penelitian

Tahapan penelitian menjelaskan proses pembangunan sistem rekomendasi kursus coding mulai dari pengumpulan data hingga evaluasi hasil rekomendasi.



Gambar 1. Flowchart Alur Proses

Tahap awal dalam pelaksanaan studi ini difokuskan pada penghimpunan data mengenai ragam kelas pemrograman di ruang lingkup *web development*, seperti yang diilustrasikan pada Gambar 1. Data yang dikumpulkan tersebut mencakup rincian topik materi, tingkatan pembelajaran, serta tautan menuju halaman kursus. Setelah dikumpulkan, data terlebih dahulu melalui proses pembersihan dan penyiapan data untuk menyederhanakan judul kursus menjadi kata-kata inti yang merepresentasikan materi pembelajaran. Hasil preprocessing tersebut selanjutnya diubah menjadi representasi numerik menggunakan TF-

IDF agar setiap term memiliki bobot sesuai tingkat kepentingannya. Setelah bobot TF-IDF diperoleh, metode *Cosine Similarity* untuk menentukan tingkat kesamaan antara masukan pengguna dengan data kursus. Nilai similarity yang dihasilkan digunakan dalam penerapan Content-Based Filtering untuk mengurutkan kursus berdasarkan tingkat kesesuaian tertinggi. Hasil rekomendasi kemudian dievaluasi menggunakan parameter Accuracy, Precision, serta Recall guna mengukur performa rekomendasi yang diberikan sistem.

### C. Sistem Rekomendasi

Sistem rekomendasi merupakan suatu pendekatan yang dirancang untuk membantu pengguna memperoleh saran terkait produk, layanan, maupun konten sesuai dengan kebutuhan, minat, dan pola perilaku yang dimiliki pengguna. Dalam implementasinya, sistem ini terdiri atas sejumlah komponen yang saling terintegrasi untuk menghasilkan rekomendasi yang tepat sasaran. Tujuan utama dari sistem rekomendasi adalah mempermudah pengguna dalam memilih alternatif yang paling relevan dari berbagai pilihan yang tersedia. Saat ini, teknologi rekomendasi telah banyak dimanfaatkan pada berbagai sektor, seperti perdagangan elektronik, layanan hiburan digital, media sosial, hingga perpustakaan berbasis digital, karena mampu meningkatkan kualitas pengalaman pengguna sekaligus mendorong keterlibatan pengguna terhadap layanan yang disediakan [9].

### D. Content-Based Filtering

*Content-Based Filtering* adalah salah satu metode pada sistem rekomendasi yang menghasilkan saran berdasarkan tingkat kesesuaian antara karakteristik suatu item dengan preferensi pengguna. Metode ini menggunakan atribut atau fitur yang dimiliki oleh item sebagai dasar dalam proses pencocokan sehingga rekomendasi diberikan kepada pengguna yang memiliki ketertarikan terhadap item dengan karakteristik serupa. Pendekatan ini banyak diterapkan pada sistem rekomendasi berbasis teks karena mampu memanfaatkan informasi yang terkandung

dalam konten tanpa memerlukan data atau penilaian dari pengguna lain [10].

### E. Metode Pembobotan TF-IDF

*Term Frequency-Inverse Document Frequency* merupakan teknik yang digunakan untuk memberikan nilai bobot pada setiap istilah dalam kumpulan dokumen teks. Besarnya bobot ditentukan berdasarkan tingkat kemunculan suatu kata pada dokumen tertentu serta jumlah dokumen lain yang juga mengandung kata tersebut. Suatu istilah akan memperoleh nilai yang tinggi apabila sering muncul pada dokumen yang sedang dianalisis namun jarang ditemukan pada dokumen lainnya. Oleh karena itu, metode ini mampu menonjolkan kata-kata yang dianggap paling mewakili isi dokumen. Pada penelitian yang menggunakan sistem rekomendasi berbasis konten, TF-IDF dimanfaatkan untuk menghasilkan representasi fitur teks yang selanjutnya digunakan dalam proses pengukuran kemiripan antar item [10].

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

Dimana:

$$TF(t, d) = \frac{f_{t,d}}{|d|}$$

$$IDF(t) = \log \left( \frac{N}{1 + df_t} \right)$$

Keterangan:

dengan t sebagai istilah yang dianalisis, d sebagai dokumen,  $f(t,d)$  sebagai frekuensi kemunculan istilah pada dokumen,  $|d|$  sebagai jumlah kata dalam dokumen, N sebagai ukuran kumpulan dokumen, serta  $df(t)$  sebagai banyaknya dokumen yang memuat istilah tersebut.

### F. Metode Cosine Similarity

*Cosine Similarity* adalah teknik yang digunakan untuk menentukan tingkat kesamaan antara dua representasi dokumen dalam bentuk vektor. Nilai kemiripan diperoleh dari perhitungan sudut yang terbentuk antara kedua vektor tersebut. Semakin tinggi nilai yang dihasilkan, semakin besar pula tingkat kemiripan antar dokumen. Dengan demikian, *Cosine Similarity* dapat digunakan untuk membantu proses pencocokan dokumen berdasarkan kesamaan topik [11].

$$\text{Cosine Similarity}(A, B) = \frac{A \cdot B}{|A||B|}$$

Keterangan:

A merupakan representasi vektor dari masukan pengguna.

B merupakan representasi vektor dari data kursus.

|A| dan |B| masing-masing menyatakan panjang vektor A dan B.

### G. Kursus Coding

Kursus online merupakan salah satu bentuk pembelajaran daring yang memungkinkan peserta memperoleh materi secara lebih fleksibel melalui media digital. Pengembangan kursus online interaktif menunjukkan bahwa pembelajaran dapat disajikan secara terstruktur dan mendukung proses belajar mandiri melalui sistem *e-learning* [12]. Coding atau pemrograman merupakan aktivitas menyusun instruksi agar komputer dapat menjalankan perintah tertentu. Dalam konteks pembelajaran, coding tidak hanya berkaitan dengan kemampuan menulis kode, tetapi juga melatih cara berpikir logis dan terstruktur dalam menyelesaikan masalah [13]. Kursus coding dapat dipahami sebagai pembelajaran terstruktur yang berfokus pada penguasaan keterampilan pemrograman dan dapat diselenggarakan melalui media daring. Pengembangan platform pembelajaran pemrograman online juga menunjukkan bahwa media belajar coding berbasis website dapat menjadi sarana yang mendukung proses belajar secara lebih mudah diakses [14].

## III. HASIL DAN PEMBAHASAN

### A. Data Penelitian

Penelitian ini memanfaatkan dataset publik yang berisi informasi kursus online yang telah diseleksi sesuai fokus penelitian, yaitu kursus coding pada domain web development. Hasil seleksi menghasilkan 15 data kursus dengan tiga atribut utama, yaitu course title, level, dan url. Atribut course title digunakan sebagai dasar pencocokan konten, level digunakan untuk menunjukkan tingkat pembelajaran, sedangkan url digunakan untuk mengarahkan pengguna ke halaman kursus.

Tabel 1. Data Kursus

| NO | Course Title                                                 | Level       | Kode URL |
|----|--------------------------------------------------------------|-------------|----------|
| 1  | Projects in HTML5                                            | Menengah    | URL-01   |
| 2  | 1 Hour HTML                                                  | Pemula      | URL-02   |
| 3  | Advanced HTML5 Tutorial for Web Developers                   | Semua Level | URL-03   |
| 4  | 1 Hour CSS                                                   | Pemula      | URL-04   |
| 5  | Coding Made Easy: HTML & CSS For Beginners                   | Pemula      | URL-05   |
| 6  | Complete Web Development with HTML, CSS and Javascripts      | Menengah    | URL-06   |
| 7  | 1 Hour JavaScript                                            | Pemula      | URL-07   |
| 8  | Learn JavaScript for Web Development                         | Menengah    | URL-08   |
| 9  | Complete PHP Course With Bootstrap3 CMS System & Admin Panel | Semua Level | URL-09   |
| 10 | Create a PHP Contact Form in Bootstrap                       | Semua Level | URL-10   |
| 11 | PSD to Bootstrap 3 for Beginners using HTML & CSS            | Pemula      | URL-11   |
| 12 | Bootstrap 4 Beginners: Code a Responsive Landing Page        | Semua Level | URL-12   |
| 13 | Create a Responsive Website with Bootstrap 3                 | Menengah    | URL-13   |
| 14 | Bootstrap 4 Quick Start: Code Modern Responsive Websites     | Semua Level | URL-14   |
| 15 | Projects in Laravel: Learn Laravel Building 10 Projects      | Menengah    | URL-15   |

Berdasarkan Tabel 1, data penelitian memuat berbagai judul kursus yang masih berada dalam ruang lingkup *web development* dengan variasi level pembelajaran yang terdiri atas Pemula, Menengah, dan Semua Level. Variasi tersebut menunjukkan bahwa data yang digunakan tidak hanya berfokus pada satu tingkat pembelajaran, tetapi mencakup beberapa tingkat yang dapat mendukung proses rekomendasi sesuai kebutuhan pengguna.

### B. Hasil Preprocessing Data

Tahap *preprocessing* dilakukan pada atribut *course title* untuk menghasilkan data teks yang lebih konsisten dan siap digunakan pada proses analisis berikutnya. Pada tahap ini, judul kursus disederhanakan dengan menghilangkan istilah yang tidak memberikan kontribusi signifikan terhadap identifikasi topik pembelajaran, sementara istilah yang memiliki keterkaitan langsung dengan materi kursus tetap dipertahankan sebagai representasi utama dari konten yang dibahas.

Tabel 2. Hasil Preprocessing Data

| NO | Course Title                        | Level       | Kode URL |
|----|-------------------------------------|-------------|----------|
| 1  | projects html                       | Menengah    | URL-01   |
| 2  | html                                | Pemula      | URL-02   |
| 3  | tutorial html web developers        | Semua Level | URL-03   |
| 4  | css                                 | Pemula      | URL-04   |
| 5  | html css                            | Pemula      | URL-05   |
| 6  | web development html css javascript | Menengah    | URL-06   |

|    |                                      |             |        |
|----|--------------------------------------|-------------|--------|
| 7  | javascript                           | Pemula      | URL-07 |
| 8  | javascript web development           | Menengah    | URL-08 |
| 9  | php bootstrap cms system admin panel | Semua Level | URL-09 |
| 10 | php contact form bootstrap           | Semua Level | URL-10 |
| 11 | psd bootstrap html css               | Pemula      | URL-11 |
| 12 | bootstrap responsive landing page    | Semua Level | URL-12 |
| 13 | responsive website bootstrap         | Menengah    | URL-13 |
| 14 | bootstrap responsive websites        | Semua Level | URL-14 |
| 15 | laravel building projects            | Menengah    | URL-15 |

Berdasarkan Tabel 2, preprocessing menghasilkan bentuk teks yang lebih sederhana tanpa menghilangkan informasi pokok dari setiap judul kursus. Kata-kata yang dipertahankan merupakan istilah utama yang berkaitan langsung dengan materi kursus, seperti *html*, *css*, *javascript*, *php*, *bootstrap*, dan *laravel*. Dengan bentuk teks yang lebih seragam tersebut, proses ekstraksi fitur pada tahap berikutnya dapat dilakukan dengan lebih terarah sehingga perhitungan kemiripan antar dokumen menjadi lebih mudah dilakukan.

### C. Hasil Pembobotan TF-IDF

Pada tahapan ini, setiap term diberikan bobot menggunakan metode TF-IDF terhadap data yang telah melewati proses preprocessing. Metode TF-IDF digunakan untuk menentukan bobot masing-masing term pada judul kursus berdasarkan frekuensi kemunculannya pada dokumen serta distribusinya pada keseluruhan dokumen. Pada penelitian ini, atribut level dimanfaatkan sebagai informasi pendukung, sedangkan atribut url digunakan sebagai tautan menuju halaman kursus. Oleh sebab itu, proses pembobotan TF-IDF difokuskan pada atribut course title yang telah melalui tahap preprocessing. Untuk mempermudah penyajian hasil, Tabel 3 hanya menampilkan sebagian dokumen fitur, sedangkan perhitungan DF dan IDF tetap dilakukan menggunakan seluruh 15 data penelitian.

Tabel 3. Perhitungan Manual

| NO | Course Title                               | Dokumen Fitur                |
|----|--------------------------------------------|------------------------------|
| 1  | Projects in HTML5                          | projects html                |
| 2  | 1 Hour HTML                                | html                         |
| 3  | Advanced HTML5 Tutorial for Web Developers | tutorial html web developers |
| 4  | 1 Hour CSS                                 | css                          |
| 5  | Coding Made Easy: HTML & CSS For Beginners | html css                     |

Untuk memperjelas pembobotan TF-IDF, digunakan salah satu dokumen fitur sebagai contoh perhitungan manual. Pemilihan dokumen

fitur html css dilakukan karena sederhana dan sesuai dengan simulasi input pengguna pada tahap Cosine Similarity. Sebagai contoh digunakan data Coding Made Easy: HTML & CSS For Beginners dengan dokumen fitur html css.

Jumlah seluruh term pada dokumen tersebut adalah 2, yaitu *html* dan *css*. Maka nilai TF dihitung sebagai berikut:

$$TF(html) = \frac{1}{2} = 0,500$$

$$TF(css) = \frac{1}{2} = 0,500$$

Selanjutnya dilakukan perhitungan IDF. Jumlah seluruh dokumen adalah 15 data kursus. Berdasarkan data *preprocessing*, term html muncul pada 6 dokumen, sedangkan term css muncul pada 4 dokumen. Maka nilai IDF dihitung sebagai berikut:

$$IDF(html) = \log\left(\frac{15}{6}\right) = \log(2,5) = 0,398$$

$$IDF(css) = \log\left(\frac{15}{4}\right) = \log(3,75) = 0,574$$

Berdasarkan hasil perhitungan pada tahap sebelumnya, diperoleh bobot akhir untuk setiap term sebagaimana ditunjukkan berikut ini:

$$TF - IDF(html) = 0,500 \times 0,398 = 0,199$$

$$TF - IDF(css) = 0,500 \times 0,574 = 0,287$$

Tabel 4. Hasil Perhitungan Manual

| Term | TF    | DF | IDF   | TF-IDF |
|------|-------|----|-------|--------|
| html | 0,500 | 6  | 0,398 | 0,199  |
| css  | 0,500 | 4  | 0,574 | 0,287  |

Berdasarkan Tabel 4, term css memiliki bobot TF-IDF lebih tinggi sebesar 0,287 dibandingkan term html sebesar 0,199. Hal ini menunjukkan bahwa term css menjadi fitur yang lebih dominan dalam merepresentasikan dokumen tersebut karena kemunculannya pada keseluruhan data lebih sedikit dibandingkan term html, sehingga nilai IDF yang dihasilkan lebih besar.

Setelah dilakukan contoh perhitungan manual, pembobotan TF-IDF diterapkan pada seluruh 15 data penelitian. Untuk mempermudah

interpretasi hasil, nilai TF-IDF pada setiap dokumen diringkas dalam bentuk fitur dominan yang difokuskan pada term-term utama yang relevan dengan domain *web development*.

Hasil pembobotan menunjukkan bahwa setiap judul kursus memiliki term dengan bobot yang lebih dominan dibandingkan term lainnya. Term-term tersebut merepresentasikan topik utama pada bidang *web development*, seperti *html*, *css*, *javascript*, *php*, *bootstrap*, dan *laravel*. Temuan ini menunjukkan bahwa metode TF-IDF mampu mengekstraksi informasi penting dari setiap dokumen sehingga representasi konten yang dihasilkan dapat dimanfaatkan pada proses pengukuran tingkat kemiripan menggunakan metode *Cosine Similarity*.

#### D. Hasil Perhitungan Cosine Similarity

Pengujian dilakukan menggunakan masukan pengguna berupa *html css* untuk mengetahui kursus yang paling sesuai dengan kebutuhan pengguna.

Tabel 5. Hasil Cosine Similarity

| NO | Course Title                                            | Similarity |
|----|---------------------------------------------------------|------------|
| 1  | Projects in HTML5                                       | 0,236      |
| 2  | 1 Hour HTML                                             | 0,570      |
| 3  | Advanced HTML5 Tutorial for Web Developers              | 0,123      |
| 4  | 1 Hour CSS                                              | 0,822      |
| 5  | Coding Made Easy: HTML & CSS For Beginners              | 1,000      |
| 6  | Complete Web Development with HTML, CSS and Javascripts | 0,468      |
| 7  | PSD to Bootstrap 3 for Beginners using HTML & CSS       | 0,490      |

Berdasarkan Tabel 5, terlihat bahwa nilai similarity tertinggi diperoleh oleh kursus Coding Made Easy: HTML & CSS For Beginners dengan nilai sebesar 1,000. Hal ini menunjukkan bahwa kursus tersebut memiliki karakteristik yang paling sesuai dengan input pengguna berupa *html css*. Kursus lain seperti 1 Hour CSS dan 1 Hour HTML juga memiliki nilai similarity yang cukup tinggi karena memiliki kesamaan term dengan input pengguna. Sementara itu, kursus dengan nilai similarity lebih rendah menunjukkan bahwa kursus tersebut masih memiliki keterkaitan dengan input pengguna, tetapi tingkat kemiripannya tidak sebesar kursus pada peringkat teratas.

#### E. Hasil Rekomendasi

Berdasarkan nilai similarity yang telah diperoleh, sistem kemudian mengurutkan kursus dari nilai kemiripan tertinggi hingga terendah. Hasil pengurutan tersebut digunakan sebagai dasar dalam menentukan rekomendasi kursus

yang paling sesuai dengan input pengguna. Pada simulasi ini, sistem menampilkan kursus dengan nilai similarity lebih dari 0, karena kursus tersebut masih memiliki kesamaan term dengan input pengguna berupa *html css*.

Tabel 6. Ranking Rekomendasi

| NO | Course Title                                            | Similarity |
|----|---------------------------------------------------------|------------|
| 1  | Coding Made Easy: HTML & CSS For Beginners              | 1,000      |
| 2  | 1 Hour CSS                                              | 0,822      |
| 3  | 1 Hour HTML                                             | 0,570      |
| 4  | PSD to Bootstrap 3 for Beginners using HTML & CSS       | 0,490      |
| 5  | Complete Web Development with HTML, CSS and Javascripts | 0,468      |
| 6  | Projects in HTML5                                       | 0,236      |
| 7  | Advanced HTML5 Tutorial for Web Developers              | 0,123      |

Berdasarkan Tabel 6, kursus Coding Made Easy: HTML & CSS For Beginners menempati peringkat pertama sebagai rekomendasi utama karena memiliki nilai similarity tertinggi sebesar 1,000. Nilai tersebut menunjukkan bahwa kursus tersebut memiliki kesesuaian paling tinggi dengan input pengguna *html css*. Kursus pada peringkat berikutnya juga masih berkaitan dengan materi HTML dan CSS, namun nilai similarity-nya lebih rendah karena hanya memiliki sebagian term yang sama atau memiliki tambahan term lain pada dokumen fitur. Dengan demikian, sistem mampu mengurutkan kursus berdasarkan tingkat kemiripan konten, sehingga pengguna dapat melihat kursus yang paling relevan hingga yang tingkat kemiripannya lebih rendah.

#### F. Pembahasan

Berdasarkan keseluruhan hasil yang diperoleh, pendekatan Content-Based Filtering mampu menghasilkan rekomendasi kursus coding dengan memanfaatkan kesamaan karakteristik pada judul kursus. Proses preprocessing berperan dalam menyederhanakan judul kursus menjadi kata-kata inti, sedangkan TF-IDF berperan dalam menentukan tingkat kepentingan masing-masing term. Selanjutnya, Cosine Similarity dimanfaatkan untuk menentukan nilai kesamaan antara masukan pengguna dan data kursus.

Pada simulasi input *html css*, sistem mampu menempatkan kursus Coding Made Easy: HTML & CSS For Beginners sebagai rekomendasi utama karena memiliki nilai similarity tertinggi. Kursus lain yang masih memiliki term *html* atau *css* juga muncul dalam

hasil rekomendasi, namun dengan nilai similarity yang lebih rendah karena hanya memiliki sebagian kesamaan atau memiliki tambahan term lain pada dokumen fitur.

Tabel 7. Hasil Evaluasi

| Parameter | Nilai  |
|-----------|--------|
| Accuracy  | 88,00% |
| Precision | 86,67% |
| Recall    | 78,57% |

Selain melakukan pengukuran kemiripan menggunakan Cosine Similarity, evaluasi kinerja sistem juga dilakukan menggunakan tiga parameter penilaian. Berdasarkan Tabel 7, diperoleh nilai Accuracy sebesar 88,00%, Precision mencapai 86,67%, dan Recall berada pada angka 78,57%. Hasil tersebut menunjukkan bahwa sistem mampu memberikan rekomendasi yang sesuai dengan masukan pengguna. Proses rekomendasi pada penelitian ini lebih menitikberatkan pada atribut *course title*, sedangkan atribut *level* dan *url* digunakan sebagai informasi tambahan.

#### IV. KESIMPULAN

Penelitian ini menunjukkan bahwa pendekatan *Content-Based Filtering* yang dipadukan dengan TF-IDF dan *Cosine Similarity* dapat dimanfaatkan untuk membangun sistem rekomendasi kursus coding berdasarkan kesamaan konten pada judul kursus. Atribut *course title* berperan sebagai sumber utama dalam pembentukan representasi konten, sedangkan atribut *level* dan *url* digunakan sebagai informasi tambahan pada hasil rekomendasi. Berdasarkan hasil pengujian, sistem mampu menentukan tingkat kesesuaian antara masukan pengguna dan data kursus, kemudian menyusun urutan rekomendasi berdasarkan nilai *similarity* yang diperoleh. Dengan demikian, sistem dapat menyajikan kursus yang memiliki relevansi paling tinggi terhadap kata kunci yang diberikan oleh pengguna.

#### UCAPAN TERIMA KASIH

Penulis menyampaikan penghargaan kepada dosen pembimbing atas arahan dan masukan selama penelitian. Apresiasi juga

diberikan kepada keluarga dan rekan-rekan yang telah memberikan dukungan serta motivasi selama proses penyusunan penelitian. Terima kasih disampaikan kepada Universitas Duta Bangsa Surakarta, khususnya Fakultas Ilmu Komputer, atas dukungan yang diberikan. Penulis turut mengapresiasi panitia (SENATIB) 2026 atas kesempatan yang diberikan untuk mempublikasikan hasil penelitian ini.

#### REFERENSI

- [1] A. A. Anggara and A. Ridho, "Sistem Rekomendasi Pembelajaran dengan Menggunakan Metode Content-Based Filtering Berbasis Aplikasi Android," pp. 11–19, 2024.
- [2] I. K. Phan, U. Widyia, and D. Pontianak, "REKOMENDASI DRAMA KOREA MENGGUNAKAN CONTENT-BASED FILTERING DENGAN WEIGHTED," vol. 13, no. 1, 2025.
- [3] S. Rahmadhani, L. Hakim, G. H. Wibowo, and S. P. Kristanto, "SISTEM REKOMENDASI PENELUSURAN BUKU BERBASIS CONTENT-BASED FILTERING DENGAN PEMBOBOTAN TF-RF," pp. 491–500, 2024.
- [4] D. Pratiwi and L. Rosnita, "Penerapan Metode Content-Based Filtering dalam Sistem Rekomendasi Objek Wisata di Aceh Tamiang," vol. 4, no. 2, pp. 85–96, 2024.
- [5] R. Faurina, E. Sitanggang, P. S. Informatika, F. Teknik, U. Bengkulu, and K. Limun, "Implementasi Metode Content-Based Filtering dan Collaborative Filtering pada Sistem Rekomendasi Wisata di Bali," vol. 22, no. 4, pp. 870–881, 2023.
- [6] H. S. Ziliwu, "DENGAN ALGORITMA COLLABORATIVE FILTERING DAN," vol. 02, pp. 141–147, 2025.
- [7] S. A. Pratama, "Pengembangan Sistem Rekomendasi Buku Menggunakan Collaborative Filtering Development Of A Book Recommendation System Using Collaborative Filtering," vol. 2, no. 2, pp. 81–86, 2024.
- [8] F. A. Hidayat and L. Fitriani, "Aplikasi Mentor Pembelajaran Berbasis Sistem Rekomendasi Content- Based Filtering dengan Metode TF-IDF dan Cosine Similarity," pp. 1048–1059, 2025, doi: 10.33364/algoritma/v.22-2.2493.
- [9] V. S. Saputra, A. Ridwan, T. G. Pratama, K. Kudus, and J. Tengah, "RANCANG BANGUN SISTEM REKOMENDASI BUKU BERBASIS ITEM-BASED COLLABORATIVE FILTERING MENGGUNAKAN ALGORITMA K-NEAREST," vol. 13, no. 1, pp. 1540–1545, 2025.
- [10] A. D. Safitri, V. Atina, A. Farida, and A. D. Safitri, "Sistem rekomendasi buku menggunakan metode content-based filtering The book recommended system employs the method of content-based filtering," vol. 5, pp. 218–227, 2024, doi: 10.37373/infotech.v5i2.1302.
- [11] K. Anam, D. H. Gutama, D. Danianti, W. D. Prastowo, F. Komputer, and U. A. Ata, "Implementasi Metode Content Based Filtering pada Pengembangan Sistem Bursa Kerja Khusus ( BKK ) SMK Negeri 1 Kandeman Berbasis Website," vol. 7, no. 4, pp. 724–733, 2024.
- [12] D. Susanto, S. Nabhan, I. Sumarsono, P. Elektronika, and N. Surabaya, "PENGEMBANGAN KURSUS ONLINE INTERAKTIF DENGAN SISTEM E-LEARNING DAN GAMIFIKASI," vol. 15, no. 2, pp. 126–137, 2022.
- [13] J. Penelitian and I. Pendidikan, "Pembelajaran Coding di Sekolah Dasar," vol. 4, pp. 62–68, 2025.
- [14] R. M. Daniswara and A. Voutama, "PERANCANGAN PLATFORM KREATIKODE SEBAGAI MEDIA PEMBELAJARAN PEMROGRAMAN ONLINE GRATIS BERBASIS WEBSITE," vol. 13, no. 2, 2025.