



Optimalisasi Literasi Digital dalam Rekomendasi Buku Referensi dengan Metode *Clustering*

Ulaya Ahdiani¹, Sri Winiarti^{2*}, Taufiq Ismail³

¹Fakultas Sastra Budaya dan Komunikasi, Sastra Inggris, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

^{2,3}Fakultas Teknologi Industri, Informatika, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
Email: ulaya.ahdiani@enlitera.uad.ac.id¹, sri.winiarti@tif.uad.ac.id², taufiq@tif.uad.ac.id³

Abstract

In the learning process, library sources or teaching references are one of the main components needed for a lecturer. The large number of textbook references available for a course can make it difficult for a lecturer to determine which option is appropriate or widely used by lecturers for the same course. The importance of selecting textbook references allows a lecturer to search for information to determine whether the reference is the main reference or an additional reference. Digital literacy is one way that a lecturer can use to search for information related to textbook references by using computers or smartphones. The problem is that the available digital literacy is not yet able to provide recommendations for textbooks that teachers widely use for a course based on certain criteria. Several criteria can be used as determining indicators for a recommended book, for example, popularity due to the highest rating used, author's name, publisher's name, content or material, and year of publication are usually the determining indicators for using a book as a reference. In fact, in cyberspace, a lot of textbook reference information is provided, but it has not been grouped based on use. By utilizing a database of books published or sold online, you can find out what textbooks are the most popular and can be used as teaching materials. Still, a lecturer needs to find out one by one in each available online shop. The method in this research is data mining for textbooks that can be used as a reference for a lecturer using text mining techniques by clustering book titles in several online stores and based on semester learning plans from several campuses to map textbooks that are widely used. as reference material for 500 book titles from several publishers. Using 6 variables consisting of a book title, use for courses or fields of science, publisher, author, year of publication, and material, analysis will be carried out to obtain a data pattern. This research aims to apply a text mining concept to provide textbook recommendations for lecturers in the Informatics field of artificial intelligence courses using clustering techniques so that they can recommend reading books using 6 variables. The aim of this research is achieved in the form of software as a form of application of theory that has been tested in the laboratory for accuracy. The benefit of this research is providing lecturers with a list of textbooks that have been ranked automatically and are widely used in the same field of science.

Keywords: Artificial Intelligence; Clustering; Cosine Similarity; Digital Literacy; College References; Recommendations

Abstrak

Dalam proses pembelajaran sumber pustaka atau referensi ajar merupakan salah satu komponen utama yang diperlukan bagi seorang dosen. Banyaknya referensi buku ajar yang tersedia untuk suatu mata kuliah dapat menjadikan seorang dosen sulit menentukan mana pilihan yang sesuai atau banyak digunakan oleh dosen untuk mata kuliah yang sama. Pentingnya dalam pemilihan referensi buku ajar memungkinkan seorang dosen melakukan pencarian informasi untuk menentukan referensi tersebut sebagai referensi utama atau sebagai referensi tambahan. Literasi digital sebagai salah satu cara yang dapat digunakan oleh seorang dosen dalam mencari informasi terkait referensi buku ajar dengan memanfaatkan komputer maupun seluler pintar. Permasalahannya adalah literasi digital yang tersedia belum mampu memberikan rekomendasi buku ajar yang banyak digunakan oleh pengajar untuk suatu mata kuliah berdasarkan kriteria tertentu. Ada beberapa kriteria yang bisa dijadikan indikator penentu suatu buku yang direkomendasikan, misalnya popularitas karena rating



tertinggi yang digunakan, nama penulis, nama penerbit, kontens atau materi, dan tahun terbit biasanya menjadi indikator penentu untuk menjadikan suatu buku sebagai referensi. Sebenarnya dalam dunia maya sudah disediakan banyak informasi referensi buku ajar, hanya belum dikelompokkan berdasarkan penggunaan. Dengan memanfaatkan basis data buku yang diterbitkan atau yang dijual secara online dapat diketahui buku ajar apa yang paling populer dan dapat dijadikan sebagai bahan ajar, namun seorang dosen perlu mencari tahu secara satu persatu pada tiap toko online yang tersedia. Metode dalam penelitian ini melakukan penambangan data buku ajar yang dapat dijadikan refensi seorang dosen dengan menggunakan teknik Text mining dengan melakukan clustring terhadap judul-judul buku yang ada di beberapa toko online dan berbasis pada Rencana pembelajaran Semester dari beberapa kampus untuk memetakan buku ajar yang banyak digunakan sebagai bahan referensi sebanyak 500 judul buku dari beberapa penerbit. Dengan menggunakan 6 variabel terdiri dari judul buku, penggunaan untuk mata kuliah atau bidang ilmu, penerbit, penulis, tahun terbit, dan materi akan dilakukan analisis untuk mendapatkan suatu pola data. Tujuan penelitian ini adalah menerapkan sebuah konsep text mining untuk memberikan rekomendasi buku ajar bagi dosen di Bidang Informatika mata kuliah kecerdasan buatan dengan menggunakan teknik-teknik clustering sehingga dapat merekomendasikan buku bacaan dengan menggunakan 6 variabel. Tujuan penelitian ini dicapai dalam bentuk sebuah perangkat lunak sebagai bentuk penerapan teori yang telah diuji di laboratorium untuk akurasi. Manfaat dari penelitian ini memberikan dosen daftar buku ajar yang telah diranking secara otomatis dan banyak digunakan untuk bidang ilmu yang sama.

Kata Kunci: Artificial Intelligence; Clustering; Cosine Similarity; Literasi Digital; Referensi Buku; Rekomendasi

1. PENDAHULUAN

Banyaknya masyarakat yang memanfaatkan internet untuk memperoleh sumber daya informasi memunculkan terjadinya suatu penggalian informasi yang dikenal dengan istilah literasi digital. Dengan memanfaatkan alat-alat digital untuk memperoleh dan menyebar informasi seperti komputer dan periperalnya serta seluler pintar inilah yang disebut literasi digital [1]. Literasi digital sendiri mulai populer semenjak terjadinya Covid-19 yang menyebabkan banyaknya pengguna memanfaatkan internet untuk bekerja, belajar, dan berkarya yang menjadikannya sumber daya yang tak terbatas oleh waktu dan lokasi. Meskipun masa Covid-19 sudah berlalu kebiasaan masyarakat dalam melakukan literasi digital tetap terus dilakukan bahkan meningkat. Tidak hanya para pengajar dan para pelajar saja yang memanfaatkan teknologi untuk melakukan literasi digital, namun masyarakat umum juga melakukan literasi digital. Literasi digital dapat juga dikatakan kemampuan seseorang untuk berinteraksi, berkomunikasi dan mencari informasi dengan melakukan interaksi melalui digital [2]. Dengan demikian maka para pengajar banyak melakukan literasi digital untuk berinteraksi dalam penggalian sumber belajar baik yang berupa buku, e-book, artikel maupun sumber website tertentu dengan tujuan pengayaan bahan ajar. Hal ini sebagai upaya adaptasi yang dilakukan oleh masyarakat khususnya para pengajar tidak hanya masa pandemi namun karena era digital yang mendukung tersedianya bahan ajar tersebut. Aktivitas literasi digital menjadi alternatif yang dapat membantu perubahan secara afektif dan menjadi pribadi yang adaptif yang berdampak meningkatkan ketahanan, ketegasan, kepercayaan diri, kemampuan mereka untuk berefleksi, untuk memecahkan masalah, serta kemampuan untuk bertahan (survival) dalam

menghadapi masalah [3]. Adanya internet mendukung proses literasi digital, sehingga e-library mudah diakses oleh siapapun [4].

Literasi Digital yang berkembang di masa ini merupakan upaya positif yang dapat digunakan oleh pengajar dan mahasiswa dengan kompetensi tertentu sehingga mempermudah mereka dalam mencari informasi untuk mengayaan sumber belajar [5]. Tujuan penguatan budaya literasi digital sebagai salah satu indikator pencapaian dalam bidang pendidikan dan kebudayaan sebagai strategi yang perlu diterapkan di lingkungan keluarga dan pendidikan dalam menunjang proses pembelajaran.

Mempertahankan budaya literasi digital ini juga dapat diterapkan dalam dunia pendidikan di perguruan tinggi. Tersedianya sumber belajar menyebabkan dosen dan mahasiswa dengan mudah memperoleh bahan ajar dan secara mandiri untuk mempelajarinya. Seorang dosen dalam memberikan materi kuliah tentunya mengacu pada suatu Rencana Pembelajaran Semester (RPS) yang disusun setiap tahun. Dalam RPS tersebut biasanya terdapat sumber pustaka utama dan tambahan yang dijadikan sebagai referensi bagi dosen yang wajib disampaikan kepada mahasiswa. Karena seorang dosen di perguruan tinggi memiliki kebijakan otonomi dalam menentukan pilihan buku ajar atau sumber pustaka yang didasari atas beberapa indikator, diantaranya; kesesuaian materi, kedalaman materi, mudah diperoleh, harga buku, penulis buku yang dikenal atau bahkan disukai serta tahun terbit. Indikator-indikator ini yang digunakan oleh para dosen dalam menjadikan buku sebagai bahan ajar yang dipakai dalam suatu mata kuliah yang diampu. Hal ini yang menyebabkan dosen dan mahasiswa lebih memanfaatkan literasi digital daripada datang secara langsung ke perpustakaan. Untuk itu perpustakaan telah beralih menerapkan digital library sebagai salah satu langkah untuk mendukung budaya literasi digital. Dampak literasi digital menyebabkan terjadinya transformasi pada perpustakaan dari konvensional menjadi perpustakaan modern ataupun perpustakaan digital. Perubahan meliputi tata kelola dan pada koleksi yang dimiliki seperti alih media atau alih format koleksi dari cetak menjadi digital. Sumber-sumber informasi berkembang dari koleksi buku tercetak menjadi buku elektronik atau sering disebut e-book. Sumber informasi tersebut juga tidak hanya berada di dalam perpustakaan, melainkan tersebar di dunia maya dan dapat diakses kapanpun dan dimanapun [6]. Kemampuan civitas akademik di perguruan tinggi untuk melakukan literasi digital sangat diperlukan. Apalagi informasi yang tersedia saat ini sangat beragam dan selalu berkembang. Keberagaman jenis informasi dalam berbagai format menambah daftar panjang kemampuan penggunaan dan pemanfaatan informasi. Friedman [7] mengilustrasikan perubahan ini sebagai “the world is flat” – yang merujuk pada sebuah keadaan dimana dunia tidak terbatas pada batas-batas negara dan zona waktu karena perkembangan teknologi. Perkembangan teknologi informasi telah menciptakan sebuah “ruang baru” yang bersifat artifisial dan maya, yang disebut *cyberspace* [8]. Peran literasi digital ini dapat dioptimalkan dengan melakukan penambahan sumber pustaka yang tersedia di internet dapat dikelola sebagai metode untuk mengelompokkan sumber pustaka tersebut ke dalam suatu mata kuliah tertentu yang dapat memperoleh suatu pola data sehingga diperolehnya sumber pustaka



mana yang paling banyak digunakan oleh para dosen untuk suatu mata kuliah tertentu. Variabel yang dipakai dapat menggunakan judul buku, mata kuliah, penerbit, penulis, tahun terbit, materi, harga dan rating buku tersebut. Dengan diperolehnya suatu pola data dapat diketahui buku yang banyak digunakan untuk bahan ajar.

Salah satu teknik yang dapat digunakan untuk melakukan penambangan data buku-buku dalam internet tersebut serta mengevaluasi pola datanya dengan menggunakan Teknik kategorisasi yang merupakan suatu teknik evaluasi dalam *text mining*. *Text mining* dapat diterapkan untuk penggalian data berupa teks dengan tipe data tidak terstruktur, diawali dari identifikasi, pemaknaan, pengelompokan dan evaluasi sehingga menjadi suatu informasi baru yang mudah dibaca [9] [10]. Dalam penelitian ini *text mining* diterapkan untuk menambang data teks dari suatu buku ajar berdasarkan judul buku, nama penulis, penerbit, harga, tahun terbit, nama mata kuliah dari beberapa toko buku online yang tersedia secara random seperti gramedia, graha ilmu, andi offset, springer, elsvier dan amazon. Teknik yang dipakai untuk mengelompokkan buku untuk mata kuliah menggunakan dua algoritma yaitu; *Cosine Similarity* dan *K-Nearest Neighbor (K-NN)*. Untuk mencari buku berdasarkan teks dengan membandingkan beberapa buku digunakan algoritma *Cosine similarity*, sedangkan untuk kategorisasi buku untuk dipetakan ke dalam suatu mata kuliah menggunakan algoritma K-NN. Kedua algoritma ini banyak digunakan untuk beberapa kasus pencarian dan pengukuran kemiripan suatu dokumen, seperti penerapan *Cosine similarity* untuk *text mining* dalam peneterjemahan Al Quran [11], Menerapkan *cosine similarity* untuk mengukur kesamaan teks yang bisa dipakai untuk klasifikasi, pengelompokkan dan pencarian *query* [12]. Penelitian lain yang hampir mirip studi kasusnya dengan penelitian yang diusulkan ini adalah penerapan metode *cosine similarity* dan K-NN untuk pencarian buku. Dalam penelitian yang dilakukan Samuel dan kawan-kawan tahun 2018 [13] berhasil melakukan pencarian buku dengan hasil pengujian rata-rata precision 0.75 %, recall 0.6 % dan f-measure 0.66 % digunakan 3 sinopsis buku untuk menghitung nilai precision, recall dan f-measure. Berdasarkan beberapa penelitian tersebut maka diusulkan penelitian untuk menerapkan *text mining* dalam literasi digital untuk rekomendasi bahan ajar bagi dosen dengan menerapkan teknik kategorisasi menggunakan algoritma *cosine similarity* dan K-NN.

Penelitian ini merupakan multi disiplin ilmu antara bidang sastra dengan computer yang menerapkan konsep *text mining* dalam memberikan rekomendasi buku. Dalam penelitian ini menunjukkan bahwa riset dasar yang dilakukan untuk bidang *Artificial Intelligence* dapat diterapkan dalam literasi digital yang merupakan keilmuan dalam bidang sastra. Hasil dari penelitian ini dapat diintegrasikan ke dalam pembelajaran mata kuliah literasi digital pada bidang Sastra Inggris dan bidang ilmu *artificial intelligence* (AI) dan data mining pada bidang Informatika.

Literasi digital memang menjadi trend sejak terjadi pandemi COVID-19 sebagai salah satu pilihan yang digunakan mendukung proses pembelajaran. Literasi digital masa COVID-19 menjadi populer karena adanya kebijakan

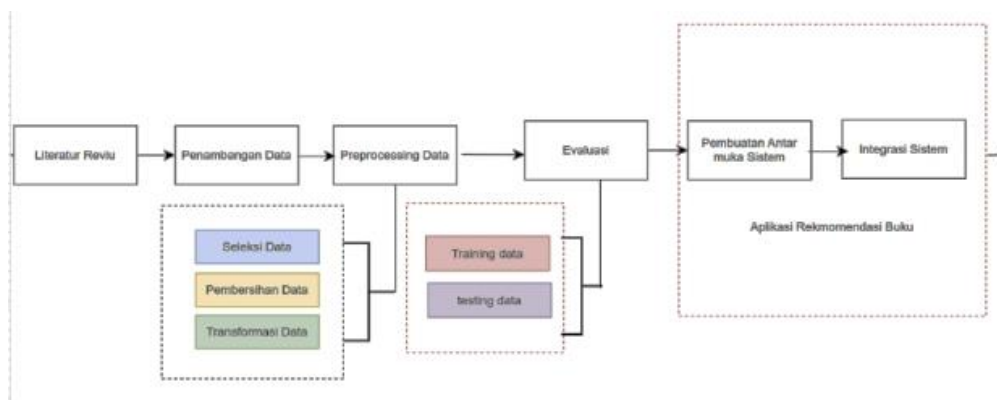
pemerintah melakukan pembatasan dalam jarak yang berdampak juga terhadap proses pembelajaran [14]. Namun dalam penelitian ini dapat menunjukkan urgensi literasi digital tidak lagi mendukung hanya pembelajaran daring, tetapi dapat dijadikan sebagai salah satu metode yang digunakan oleh pelajar dan pengajar untuk mendapatkan referensi buku ajar yang populer digunakan.

Dalam penelitian ini melakukan pengembangan literasi digital dengan memanfaatkan keilmuan *Artificial Intelligence* (AI) yang digunakan guna menambang data berupa buku-buku referensi bacaan yang rekam jejak datanya diambil dari berbagai sumber data yang ada di internet. Konsep yang digunakan adalah menggunakan *clustering* dan *cosine similarity*. *Clustering* digunakan untuk mengelompokkan data buku berdasarkan kelasnya, sedangkan *cosine similarity* digunakan untuk mengidentifikasi buku berdasarkan data teks dari judul buku yang telah ditambang. Berdasarkan penelitian yang telah menerapkan metode *clustering* dalam mengelompokkan data, algoritma K Means sebagai salah satu algoritma yang banyak digunakan. Dalam penelitian ini menggunakan algoritma K-Means karena proses pengelompokan dapat dilakukan secara berulang kali di dalamnya. Algoritma ini menggunakan sejumlah nilai *cluster* (k) yang acak dengan menjadikan nilai tersebut sebagai *centroid* atau pusat *cluster*, *means* atau rerata yang digunakan sebagai nilai yang mewakili dari sekumpulan data. Beberapa penelitian yang menggunakan algoritma K-Means sebagai proses rekomendasi karena dalam implementasinya dinilai akurat, cepat, mudah, dan sederhana untuk mengolah data, berbanding terbalik dengan algoritma *clustering* lainnya [15] [16] [17].

2. METODOLOGI PENELITIAN

2.1. Tahapan Penelitian

Dalam penelitian ini metode yang digunakan terdiri dari beberapa tahap, yang diawali dari literatur reuiu, penelusuran masalah, perumusan masalah, pengumpulan data, pengolahan data, pengembangan sistem dan pengujian. Gambar 1 merupakan tahapan dari metode riset yang dilakukan [18] [19].



Gambar 1. Tahapan Penelitian

2.2. Literatur review

Digunakan dengan mengkaji penelitian yang sudah pernah ada terkait algoritma yang digunakan untuk diterapkan dengan metode clustering. Tujuan dilakukannya literatur revidu untuk mencari gap yang terjadi dalam penelitian sehingga dapat ditemukan suatu solusi atau metode yang sesuai. Kajian terhadap literatur pada riset ini dilaksanakan dengan cara menggali referensi dan permasalahan dari sejumlah penelitian yang pernah dilakukan sebelumnya dan berhubungan dengan topik yang diangkat. Referensi yang digunakan berasal dari jurnal ilmiah, buku elektronik, serta skripsi mengenai algoritma K-Means yang didapat dari berbagai macam sumber yang ada di *internet* dengan bantuan Google Scholar.

2.3. Penambahan Data

Penambahan data dilakukan dengan cara mengambil data dari sumber pustaka yang terdapat di toko buku secara online, digilib perpustakaan UAD dengan mengambil data buku yang terdiri dari koleksi buku informatika milik Perpustakaan Nasional untuk selanjutnya diteliti menggunakan sistem rekomendasi yang memanfaatkan algoritma K-Means. Atribut *dataset* dalam penelitian ini dihimpun berdasarkan hasil kuesioner mengenai kriteria apa saja yang sering kali dipertimbangkan dalam memilih buku. Sedangkan himpunan data bukunya sendiri digali dari koleksi milik Perpustakaan Nasional, serta sejumlah *website* penerbit dan toko buku. Dari hasil pengumpulan ini, didapatkan tiga ratus *record* dengan delapan buah atribut atau kriteria buku.

2.4. Preprocessing Data

Proses pengolahan akan diawali dengan *preprocessing* yang terdiri dari seleksi, pembersihan data, dan transformasi untuk menghasilkan data yang sesuai dengan kebutuhan sebelum diolah lebih lanjut menggunakan K-Means. Selanjutnya, himpunan data akan dicari tahu jumlah *cluster* idealnya dengan mengaplikasikan rumus pada *Elbow Method* dan *Silhouette Score*. Setelah data hasil pencarian buku diperoleh dengan menggunakan *teknik cosine similarity* dilakukan klasifikasi dengan menggunakan algoritma K-Means. Klasifikasi ini diperlukan untuk mengelompokkan data buku yang paling banyak digunakan untuk mata kuliah *artificial intelligence* dan sastra inggris. Sehingga dapat sebagai rekomendasi bagi dosen dan mahasiswa dalam menentukan buku ajar untuk kedua mata kuliah tersebut. Penentuan rekomendasi buku berdasarkan pola data yang dihasilkan dari penerapan algoritma K-Means yang dilakukan melalui training dan testing. Sehingga hasil uji akurasi yang tertinggi yang dapat dijadikan rekomendasi untuk penentuan kelompok buku tersebut. Implementasi teknik ini menggunakan pemrograman Python.

2.5. Evaluasi

Keseluruhan bagian sistem akan diperiksa dan diuji guna mengetahui adanya probabilitas terjadinya kegagalan atau *error* dengan cara memasukkan data yang kemudian diolah oleh sistem untuk mendapatkan hasil ujinya.

Tahap evaluasi dilaksanakan menggunakan dua metode yaitu *Fowlkes-Mallows Index* (FMI) dan *User Acceptance Testing* (UAT). Adanya evaluasi dengan FMI bertujuan untuk mengetahui seberapa baik kualitas hasil pengelompokan menggunakan K-Means. Sedangkan UAT digunakan untuk memverifikasi kinerja sistem yang langsung digunakan *user*.

2.6. Merancang antar muka Sistem

Design adalah hasil tindak lanjut dari analisis kebutuhan yang telah dilakukan. Keseluruhan aspek kemudian dituangkan dalam bentuk lain seperti diagram aktivitas dan rancangan antarmuka (*interface*). Lantaran riset yang akan dilakukan ini memanfaatkan aplikasi dengan teknologi *website*, maka nantinya akan dibuat halaman-halaman untuk memproses dan menampilkan data buku yang diolah menggunakan algoritma K-Means.

2.7. Integrasi Sistem

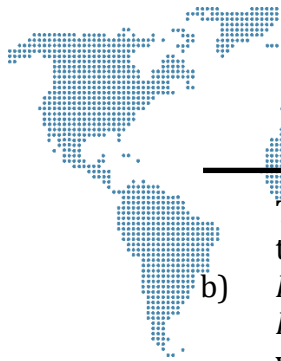
Melakukan integrasi hasil desain dan analisis ke dalam aplikasi yang diimplementasikan dengan PHP karena system yang dikembangkan dibuat menggunakan teknologi web. Tujuannya agar pengguna dapat mengoperasikan system untuk mencari referensi sesuai kebutuhannya. Rancangan antarmuka akan diterapkan ke dalam bentuk nyata berupa sistem rekomendasi dengan basis *website* yang dikembangkan melalui susunan baris kode PHP dan HTML. Guna mendukung proses implementasi ini, diperlukan pula sejumlah *tools*, dua di antaranya yaitu Visual Studio Code dan XAMPP.

3. HASIL DAN PEMBAHASAN

3.1. Perolehan Dataset

Dalam pengumpulan data yang dilakukan dari berbagai sumber pustaka secara online maka data yang diperoleh sebanyak 301 judul buku dari berbagai penerbit. Atribut *dataset* dalam penelitian ini dihimpun berdasarkan hasil kuesioner mengenai kriteria apa saja yang sering kali dipertimbangkan dalam memilih buku. Dari hasil pengumpulan ini, didapatkan tiga ratus *record* dengan delapan buah atribut atau kriteria buku. Jumlah total data yang diperoleh sebanyak 301 judul buku. Selanjutnya dilakukan pengolahan data buku dengan menggunakan tahapan dalam *text mining*, yaitu; *data selection*, *data cleaning*, dan *data transformation*. Selain *preprocessing data*, dilakukan juga pencarian jumlah kluster K-Means yang ideal. Penjelasan setiap tahapan tersebut sebagai berikut:

- a) *Data selection* adalah langkah memilih atribut apa saja yang akan dipakai dalam proses *clustering* sehingga data lain yang tidak digunakan dapat diabaikan. Dalam hal ini terdapat tiga buah kriteria yang dapat digunakan dalam proses perhitungan K-Means. Ketiga atribut dari kriteria buku yang dipilih meliputi atribut tahun, jumlah awal, jumlah keluar, dan harga. Proses *data selection* membuat kriteria penulis, penerbit, dan ISBN dihapus karena tidak digunakan dalam pengolahan yang memanfaatkan algoritma K-Means sehingga hanya menyisakan lima buah kriteria yaitu judul sebagai penanda masing-masing *record*, tahun, jumlah awal, jumlah keluar, dan harga buku.



Tabel 1 dan 2 menunjukkan data hasil sebelum dan sesudah *preprocessing* tahap pertama yakni *data selection*.

b) *Data Cleaning*

Data cleaning dilakukan untuk mengidentifikasi duplikasi atau data tidak valid. Data-data tersebut kemudian dapat dihapus atau diganti. *Dataset* buku yang telah dihimpun dalam riset ini memiliki duplikasi pada salah satu *record* yang berjudul “Buku Ajar Web Dasar”, yang terletak pada nomor 2 dan 299 sehingga salah satunya dihapus. Proses *data cleaning* dilakukan dengan menghapus *record* yang berjudul “Buku Ajar Web Dasar”, yang terletak pada nomor 299 sehingga menghasilkan 300 *record* yang siap untuk diolah system.

c) *Data Transformation*

Data transformation merupakan tahap ubah data menjadi bentuk lain dengan tujuan mudah untuk diolah. Transformasi yang dilakukan dalam riset ini berupa penggabungan jumlah awal dan jumlah keluar menjadi *rating* buku. Selain itu, terdapat atribut tahun yang akan diubah menjadi usia buku dengan cara mencari selisih tahun aktual dengan tahun terbit yang tertera pada *dataset*. Setelah melalui langkah penggabungan, seluruh *record* yang ada pada ketiga atribut atau kriteria baru tadi akan dinormalisasi dan data diubah dengan *range* satu sampai tiga menggunakan Persamaan 1 [20].

$$V_i' = \frac{V_i - \text{Min}_A}{\text{Max}_A - \text{Min}_A} \quad \text{New_max}_A - \text{New_Min}_A + \text{New_Min}_A \quad (1)$$

Tujuan menggunakan persamaan 1 mendapatkan data yang ternormalisasi agar data yang telah diubah dapat digunakan sebagai dataset yang digunakan ke dalam system literasi digital. Dengan menggunakan persamaan 1 maka data dapat diperoleh hasilnya ternormalisasi seperti yang ditunjukkan pada Tabel 1.

Tabel 1. Hasil Data Normalisasi

No	Judul buku	Penulis	Usia	Rating	Harga
1	Dasar-dasar penguasaan pemrograman	Agusriandi	1.29	2.0	1.42
2	Buku Ajar Web Dasar	Stepanus Widji	1.34	1.00	1.22
3	Pemrograman Web Dasar	Rintho Rante Rerung	1.27	1.50	1.32
...	...	...	...	...	...
298	Pemrograman dasar java visual berbasis database	Ramadhani	1.40	2.33	1.24
299	Dasar - Dasar Teknik Informatika	Novega Pratama	1.17	1.37	1.29
300	Konsep grafika komputer	Dr. Pulung Nurtantio Andono	1.40	1.90	1.26

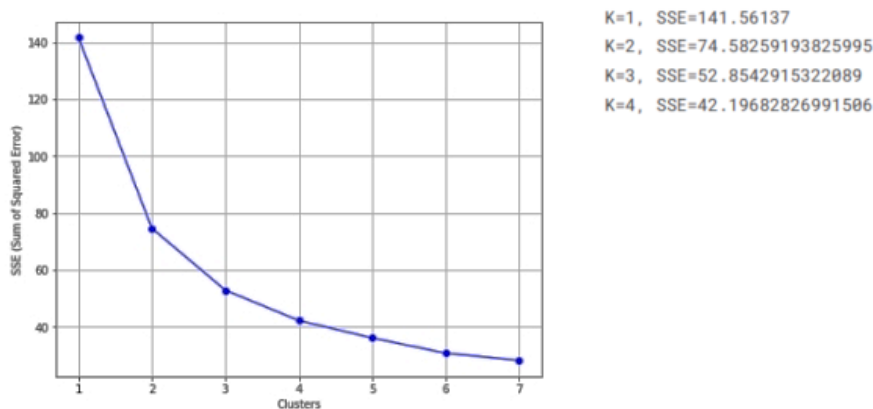
Dalam keilmuan AI tahapan *data selection*, *data cleaning*, dan *data transformation* merupakan proses penambangan data teks (data mining). Tujuan tahapan ini agar diperolehnya pengelompokkan data buku referensi berdasarkan jenis buku yang dapat dijadikan sebagai kumpulan data yang dapat dijadikan literasi digital untuk rekomendasi bahan bacaan. Dengan demikian konsep AI yang digunakan untuk membuat literasi digital referensi buku ajar.

Literasi digital yang dimaksud dalam penelitian ini berupa sebuah aplikasi yang mengoleksi berbagai sumber ajar yang dapat digunakan oleh mahasiswa dan

dosen dalam mencari referensi yang paling banyak digunakan oleh orang sebagai referensi. Gambar 2 menunjukkan tampilan awal dari system literasi digital yang menjadi produk tepat guna dalam penelitian ini.

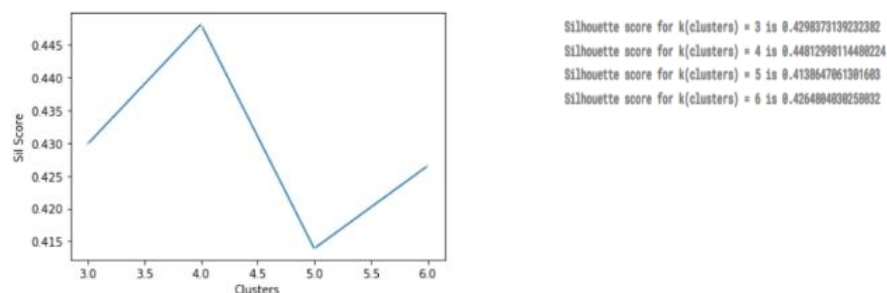
3.2. Penentuan Jumlah Cluster

Berisi hasil pembahasan dan bisa perbandingan dari hasil penelitian sebelumnya. Penentuan jumlah cluster menggunakan *Elbow Method* dengan memberikan rentang nilai 2 sampai 4 cluster. Gambar 2 menunjukkan grafik penentuan jumlah k menggunakan *Elbow Method*.



Gambar 2. Grafik visualisasi cluster dengan *Elbow Method* dan perhitungan SSE

Gambar 2 menunjukkan hasil perhitungan *Sum of Squared Errors* (SSE) dalam penentuan jumlah k menggunakan *Elbow Method*. Jika dilihat pada cluster keempat, *error* yang dihasilkan jumlahnya paling kecil atau dengan kata lain mendekati 0 sehingga cluster empat merupakan *cluster* ideal.



Gambar 3. Grafik *houette* Score dan Hasil Perhitungan SSE

Setelah menentukan nilai cluster dengan *Elbow Method* selanjutnya perlu dilakukan validasi agar jumlah k ideal yang telah dicari menggunakan *Elbow Method* sebelumnya. Hal ini dapat dilakukan dengan menggunakan *Silhouette Score*. Hasilnya ditunjukkan pada Gambar 3 yang merupakan visualisasi dari hasil *Silhouette Score*. Berdasarkan Gambar 3 dapat dilihat jika titik *cluster* empat berada di posisi tertinggi dibandingkan *cluster-cluster* lain. Semakin tinggi

Silhouette Score yang dihasilkan, maka akan semakin tinggi pula kualitas pengelompokannya. Jika hasilnya semakin mendekati 0, maka ada tumpang tindih antara *cluster* dan *centroid*, yang memungkinkan terdapat pengelompokan data yang kurang baik. Jika dilihat pada *cluster* keempat, angka yang dihasilkan jumlahnya paling besar atau dengan kata lain semakin mendekati 1 sehingga cluster empat yang merupakan *cluster* ideal.

3.3. Evaluasi

Evaluasi dilaksanakan guna mengetahui seberapa baik kinerja algoritma yang dipakai, dalam kasus ini terkait kualitas hasil pengelompokan menggunakan K-Means. Untuk mengetahui seberapa tinggi kualitasnya, maka digunakan metode *Fowlkes-Mallows Index* (FMI) untuk mengevaluasi. Model pengelompokan data menggunakan *index cluster* yang dimulai dari satu, sedangkan *script* Python menggunakan *index* yang dimulai dari nol. Hasil dari Python akan digunakan sebagai *ground-truth* atau *dataset* "kebenaran" atau referensi akurat yang membantu menentukan sejauh mana model berhasil dalam tugas klasteringnya. Gambar 4 menunjukkan proses pencocokan dan kalkulasi nilai FMI. Kalkulasi nilai FMI di atas menunjukkan hasil sebesar 1. Seperti yang telah diketahui sebelumnya, semakin nilai FMI mendekati satu, maka hasil pengelompokan yang dilakukan model akan semakin baik. Maka dari itu, dapat disimpulkan jika model berhasil melakukan tugasnya dengan baik.

```
# Pencocokan Index
mapping = {1: 0, 2: 1, 3: 2, 4: 3}
kmeans_labels = df_kmeans['Cluster'].map(mapping).tolist()

# Rumus Fowlkes-Mallows Index (FMI)
TP = sum(1 for a, b in zip(kmeans_labels, expected_labels) if a == 1 and b == 1)
FP = sum(1 for a, b in zip(kmeans_labels, expected_labels) if a == 1 and b != 1)
TN = sum(1 for a, b in zip(kmeans_labels, expected_labels) if a != 1 and b != 1)
FN = sum(1 for a, b in zip(kmeans_labels, expected_labels) if a != 1 and b == 1)

Precision = TP / (TP + FP)
Recall = TP / (TP + FN)
FMI = sqrt(Precision * Recall)

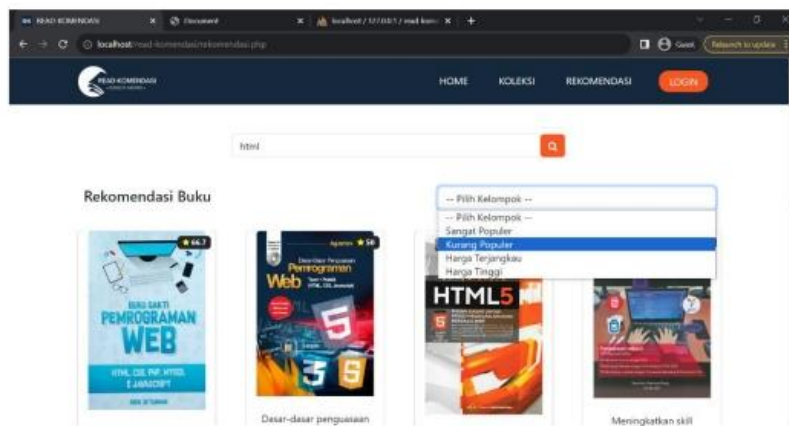
# Cetak hasil perhitungan
print("Fowlkes-Mallows Index (FMI):", FMI)
```

Fowlkes-Mallows Index (FMI): 1.0

Gambar 4. Proses Pencocokan dan kalkulasi nilai FMI

3.4. Integrasi Sistem

Setelah dilakukan analisis untuk melakukan clusteringnya, selanjutnya diintegrasikan kedalam desain antarmuka system, agar system dapat digunakan oleh pengguna dalam mencari referensi buku. Pengintegrasian Pemrograman Python dengan PHP dilakukan agar menjadi suatu system yang dapat diakses dengan teknologi web. Gambar 4 menunjukkan contoh antarmuka aplikasi rekomendasi buku referensi sebagai bentuk implementasi system.



Gambar 5. Contoh implementasi iterasi digital untuk aplikasi rekomendasi buku referensi

Metode clustering diterapkan ke dalam antarmuka system sebagai bentuk visualisasi dari aplikasi menampilkan hasil clustering buku, sedangkan pemrograman pyhton sebagai mesin yang melakukan training dan validasi untum membaca pola data. Hal ini ditunjukkan pada Gambar 6.

No	Judul Buku	Uraian	Rating	Harga	Hasil
1	Dasar-dasar penggunaan pemrograman web - teori & praktik (HTML, CSS, JAVASCRIPT)		5	10	Cluster-2
2	Buku Ajar Web Dasar		5	0	Cluster-2
3	Pemrograman Web Dasar		5	25	Cluster-2
4	Dasar Pemrograman Web Dinamis Menggunakan PHP: 800 Rival Koding		3	12,5	Cluster-1
5	Pemrograman Web Edisi Revisi		5	75	Cluster-4
6	Dasar Pemrograman Web dengan PHP		1	65,57	Cluster-4
7	HTML5 Dasar-Dasar untuk Pengembangan Aplikasi Berbasis web		4	50	Cluster-2

Gambar 6. Hasil visualisasi sebagai antarmuka system dalam clustering buku referensi

4. SIMPULAN

Berdasarkan pada penelitian yang telah dilakukan, maka dapat ditarik kesimpulan bahwa sistem mengelompokkan data menjadi empat *cluster*. Masing-masing mewakili kriteria berbeda, yaitu C1 dengan harga buku tinggi, C2 adalah klaster dengan harga terjangkau, C3 sebagai klaster kurang populer, dan terakhir C4 dengan buku-buku yang populer. Data buku yang ditampilkan oleh sistem sebagai rekomendasi pada masing-masing kelompok diurutkan berdasarkan tingkat kepopulerannya. Evaluasi dengan *Fowlkes-Mallows Index* (FMI) memberikan hasil sebesar 1 sehingga K-Means terbukti mampu menyelesaikan kasus rekomendasi buku. Aplikasi ini dapat dijadikan sebagai model literasi digital bagi mahasiswa ataupun pengajar yang ingin mencari buku referensi dan telah diujikan dengan hasil baik.

DAFTAR PUSTAKA

- [1] Sutrisna, I. P. G, "Gerakan Literasi Digital Pada Masa Pandemi Covid-19.," *Stilistik: Pendidikan Bahasa Dan Seni*, Vol. 8(2), No. <https://doi.org/10.5281/zenodo.3884420>, P. 268–283, 2020.
- [2] Ayu,I.R, "Keterkaitan Literasi Dengan Adaptasi Mahasiswa Di Masa Pandemi," *Jurnal Amal Pendidikan*, Vols. Vol. 1, No. 2, Agustus2020, No. [Doi: Http://Dx.Doi.Org/10.36709/japend.v1i2.13214](http://dx.doi.org/10.36709/japend.v1i2.13214), Pp. Hal: 122-128, , 2020.
- [3] Jones, R. H. And Hafner, C. A. , *Understanding Digital Literacies: A Practical Introduction.*, Routledge.Com, 2012.
- [4] R. Yasir, S. Mohammad, M. Mailizar, T. M. Alqahtani, L. . N. Yaqin And A. Habibie, "How Can We Assess The Success Of Information Technologies In Digital Libraries? Empirical Evidence From Indonesia," *International Journal Of Information Management Data Insights*, 2023.
- [5] Giovanni, F., And Komariah, N. , "Hubungan Antara Literasi Digital Dengan Prestasi Belajar Siswa Sma Negeri 6 Kota Bogor.," *Libraria: Jurnal Perpustakaan*, Vol. 7(1), P. 147, (2020).
- [6] Rahmawati, N.A, Faudah, L And Suryanta, B, "Peningkatan Literasi Informasi Dan Minat Baca Melalui Pelatihan Penelusuran E-Journal Dan E-Book Di Perpustakaan Upn Veteran Yogyakarta," *Jurnal Perpustakaan* , Vol. Vol. 10 No.2 , No. Issn 1979 - 9527, Pp. 103-108, Tahun 2019: .
- [7] Afandi, , Junanto, T., And Afriani, R., "Implementasi Digital-Age Literacy Dalam Pendidikan Abad 21 Di Indonesia.," In *In Prosiding Seminar Nasional Pendidikan Sains*, 2016.
- [8] Piliang, Y. A. , "Mayarakat Informasi Dan Digital: Teknologi Informasi Dan Perubahan Sosial.," *Jurnal Sosioteknologi*, Vols. 27(11),, P. 143–156., 2012.
- [9] Konchady, M., *Text Mining Application Programming (Programming Series) 1st Edition*, Charles River Media Programming, 2006.
- [10] R. Zhang , J. Zhang, . Q. Chen, Q. Qian , D. Pan, J. Xia , Y. Wang And Y. Han, "A Literature-Mining Method Of Integrating Text And Table Extraction For Materials Science Publications," *Computational Materials Science*, Vol. 230, 2023.
- [11] M. Didik And R. Wahyudi, "Penerapan Algoritma Cosine Similarity Pada Text Mining Terjemah Al-Qur'an Berdasarkan Keterkaitan Topik," *Semesta Teknik*, Vols. Vol. 22, No.1,, No. [Doi: 10.18196/st.221235](https://doi.org/10.18196/st.221235), Pp. 41-50, 2019.
- [12] Sohangir,S And Wang,D, "Improved Sqrt-Cosine Similarity," *Journal Big Data* , Vol. 4, No. [Doi 10.1186/s40537-017-0083-6](https://doi.org/10.1186/s40537-017-0083-6), (2017) .
- [13] Samuel,R, Natan, R, Fitria, And Syafiqoh,U, "Penerapan Cosine Similarity Dan K-Nearest Neighbor (K-Nn) Pada Klasifikasi Dan Pencarian Buku," *Journal Of Big Data Analytic And Artificial Intelligence*, Vols. Vol. 1, No. 1,, Pp. 9-14, Tahun 2018.
- [14] Ida Wahyu Ningsih , Arif Widodo And Asrin, "Urgensi Kompetensi Literasi Digital Dalam Pembelajaran Pada Masa," *Jurnal Inovasi Teknologi Pendidikan*, Pp. 133-139, 2021.
- [15] K. Al Atros, A. Padri, O. Nurdiawan, A. Faqih And S. Anwar, "Model Klasifikasi Analisis Kepuasan Pengguna Perpustakaan Online Menggunakan K-Means Dan Decission Tree," *Jurikom (Jurnal Riset Komputer)*, Vol. Vol. 8 No. 6, Pp. 323-329, 2021.
- [16] L. Fimawahib And E. Rouza, "Penerapan K-Means Clustering Pada Penentuan," *Jurnal*

Inovtek Polbeng, Vol. Vol 6 No 2, No. Seri Informatika, Pp. 234-247, 2021.

- [17] B. G. Aji, D. C. A. Sondawa, M. R. Gifari And S. W, "Penerapan Algoritma K-Means Untuk Clustering Harga Ruman Di Bandung," *Jurnal Ilmiah Informatika Global*, Vol. Vol. 14, 2023.
- [18] M. Bolung And H. R. K. Tampangela, "Analisa Penggunaan Metodologi Pengembangan Perangkat Lunak,," *Eltikom*, , Vol. Vol. 1, 2017.
- [19] I. R. Gumilang, "Penerapan Metode Sdlc (System Devlopment Life Cycle) Pada Website Penjualan Produk Vapor," *Jurnal Riset Rumpun Ilmu Teknik*, Vol. Vol 1, 2022.
- [20] D. Ulandari And D. Swanjaya, "Perbandingan Transformasi Data Pada Penentuan Peserta Bimbingan Belajar Menggunakan Metode Perceptron," In *Seminar Nasional Inovasi Teknologi Un Pgri*, Kediri, , 25 Juli 2020.