

IMPLEMENTASI PENDEKATAN *COLLABORATIVE FILTERING* DAN *K-MEANS CLUSTERING* PADA SISTEM REKOMENDASI MATA KULIAH

Edward Fernando¹, Panca Mudjiraharjo², Muhammad Aswin³

^{1,2,3} Teknik Elektro, Fakultas Teknik, Universitas Brawijaya
Email: ¹edward@ub.ac.id, ²panca@ub.ac.id, ³maswin@ub.ac.id

(Naskah masuk: 22 Juni 2022, diterima untuk diterbitkan: 27 Juni 2022)

Abstrak

Dalam perkuliahan, mahasiswa akan menjalani beberapa rangkaian mata kuliah, secara umum prodi akan menyiapkan daftar mata kuliah baik yang bersifat wajib dan pilihan. Prodi telah menyiapkan kurikulum yang berlaku dalam bentuk silabus, namun tidak jarang mahasiswa ragu dalam mengambil keputusan mata kuliah pilihan apa yang akan mereka ambil, terlebih lagi pada awal menentukan konsentrasi keminatan prodi. Hal ini dapat berakibat kurang maksimalnya perolehan nilai mata kuliah mahasiswa hingga mundurnya masa studi. Penelitian terdahulu telah melakukan pendekatan sistem pakar dan *data mining* untuk menentukan rekomendasi mata kuliah sesuai bakat dan kemampuan individu. Namun dalam implementasinya, tidak jarang mata kuliah yang direkomendasikan terlalu luas dan kurang sesuai dengan profil pengguna, dengan demikian perlu adanya metode optimalisasi rekomendasi untuk meningkatkan akurasi rekomendasi dengan meminimalisir keluarnya rekomendasi mata kuliah yang kurang sesuai dengan profil data uji. Dalam penelitian ini, sistem rekomendasi akan dibangun dengan pendekatan *collaborative filtering* menggunakan metode *jaccard*, *euclidean distance*, *cosine similarity*, dan *pearson correlation*. Setelah itu, *k-means clustering* digunakan untuk mengoptimalkan nilai akurasi dari sistem rekomendasi dengan . Dari penelitian yang dilakukan, pendekatan *collaborative filtering* menggunakan metode *cosine similarity* menghasilkan nilai akurasi 64.76%, *euclidean distance* menghasilkan nilai akurasi 62.04%, *jaccard* menghasilkan nilai akurasi 53.66%, dan *pearson correlation* menghasilkan nilai akurasi 43.71%. Optimalisasi akurasi rekomendasi menggunakan *k-means clustering* juga berhasil meningkatkan performa nilai akurasi hingga 9.92%.

Kata kunci: *collaborative filtering*, *k-means clustering*, mata kuliah, sistem rekomendasi.

IMPLEMENTATION OF *COLLABORATIVE FILTERING* AND *K-MEANS CLUSTERING* APPROACHES IN THE COURSE RECOMMENDATION SYSTEM

Abstract

In lectures, students will undergo several series of courses, in general the study program will prepare a list of both mandatory and optional courses. The study program has prepared an applicable curriculum in the form of a syllabus, but it is not uncommon for students to hesitate in making decisions about which elective courses they will take, especially at the beginning of determining the concentration of study program interests. This can result in less than the maximum value of student course scores until the study period is delayed. Previous research has carried out an expert system approach and data mining to determine course recommendations according to individual talents and abilities. However, in its implementation, it is not uncommon for the recommended courses to be too broad and not in accordance with the user profile, thus it is necessary to have a recommendation optimization method to increase the accuracy of the recommendations by minimizing the issuance of course recommendations that are not in accordance with the test data profile. In this study, a recommendation system will be built using a collaborative filtering approach using the jaccard, euclidean distance, cosine similarity, and pearson correlation methods. After that, k-means clustering is used to optimize the accuracy value of the recommendation system. From the research conducted, the collaborative filtering approach using the cosine similarity method produces an accuracy value of 64.76%, euclidean distance produces an accuracy value of 62.04%, jaccard produces an accuracy value of 53.66%, and pearson correlation produces an accuracy value of 43.71%. Optimization of recommendation accuracy using k-means clustering also succeeded in increasing the performance of the accuracy value up to 9.92%.

Keywords: *collaborative filtering*, course, *k-means clustering*, recommendation system.

1. PENDAHULUAN

Dalam sebuah perkuliahan, mahasiswa akan menjalani beberapa rangkaian mata kuliah yang telah disiapkan oleh program studi untuk mencapai gelar sarjana. Secara umum, program studi akan menyiapkan daftar mata kuliah yang ditawarkan sebagai gambaran kepada mahasiswa mengenai hal-hal apa saja yang akan mereka pelajari. Mata kuliah pilihan yang mahasiswa pilih sering kali akan menentukan keminatan apa yang akan ditekuninya. Oleh karena itu, program studi menerbitkan silabus kurikulum untuk membantu mahasiswa dalam memperkenalkan pokok bahasan tiap mata kuliah secara umum beserta tuntunan kapan seharusnya mahasiswa mengambil mata kuliah tersebut. Sering kali mahasiswa merasa kebingungan untuk memilih mata kuliah pilihan yang dibebaskan oleh kurikulum. Untuk membantu memilih mata kuliah pilihan yang sesuai, mahasiswa dapat meminta pertimbangan dosen penasihat akademik. Namun terkadang penasihat akademik tidak dapat membantu banyak dalam memberi pertimbangan, mengingat mahasiswa bimbingan mereka yang cukup banyak serta keterbatasan waktu yang ada, sehingga banyak mahasiswa yang mengambil keputusan untuk mengambil mata kuliah pilihan tanpa pertimbangan yang matang.

Dalam membantu mahasiswa untuk menentukan mata kuliah pilihan mana yang sesuai dengan bakat dan kemampuan masing-masing individu, terdapat beberapa pendekatan yang dapat dilakukan seperti penggunaan sistem pakar [1] dan juga *data mining* [2]. Akhir-akhir ini, pengkajian *Educational Data Mining* (EDM) cukup sering dilakukan. Pada penelitian sebelumnya, EDM digunakan untuk melakukan profiling performa mahasiswa guna mengetahui apa saja faktor yang mempengaruhi performa akademis mahasiswa [3]. EDM juga telah berhasil untuk memberi rekomendasi pemilihan topik tugas akhir [4]. Selain itu, EDM juga mampu mengidentifikasi mata kuliah apa saja yang mempengaruhi mahasiswa *drop out* atau mengganti keminatannya [5]. Tidak hanya sekedar mengidentifikasi dan menganalisa performa akademis mahasiswa, EDM juga dapat digunakan untuk membantu mahasiswa dalam memilih mata kuliah pilihan [6].

Collaborative filtering secara umum adalah salah satu pendekatan yang dapat digunakan sebagai metode sistem rekomendasi berdasarkan suatu anggapan bahwa seseorang akan memiliki ketertarikan yang sama dengan pengguna yang memiliki preferensi yang sama terhadap suatu hal [7]. *Collaborative filtering* merupakan pendekatan yang sering kali digunakan dalam kasus mengolah banyak pilihan menjadi bentuk yang lebih sederhana dengan mengolah penilaian interaksi seseorang terhadap pilihan yang tersedia. Dengan demikian,

pengguna baru dapat disarankan untuk menikmati konten yang telah dinikmati oleh pengguna yang telah memiliki kemiripan terbanyak dengannya karena asumsi dasarnya mereka memiliki ketertarikan yang sama pada suatu genre, cerita, atau tokoh yang berperan. *Collaborative filtering* akan mempelajari data yang ada dan mencocokkan dengan profil mahasiswa yang sedang memilih mata kuliah pilihan. Metode ini akan mempelajari riwayat mata kuliah yang telah dipelajari oleh mahasiswa-mahasiswa terdahulu untuk mempelajari bagaimana mahasiswa terdahulu berinteraksi dengan mata kuliah yang telah mereka jalani selama menempuh perkuliahan. *Collaborative filtering* akan mempelajari data yang ada dan mencocokkan dengan profil mahasiswa yang sedang memilih mata kuliah pilihan. Apabila terdapat kemiripan profil, *collaborative filtering* akan menyajikan daftar mata kuliah pilihan yang dapat dipilih sebagai rekomendasi yang paling sesuai.

Pada penelitian sebelumnya, *collaborative filtering* terbukti mampu untuk merekomendasikan mata kuliah pilihan kepada mahasiswa [6]. Pada penelitian tersebut, pendekatan yang digunakan pada *collaborative filtering* adalah *Item-Based* dan *User-Based*. Pada penelitian lain, *collaborative filtering* juga dapat diterapkan untuk memprediksi IPK mahasiswa [8]. Pada implementasi sistem rekomendasi menggunakan pendekatan *collaborative filtering*, hasil rekomendasi umumnya hanya serta merta menampilkan *item-item* yang dimiliki oleh data latih yang memiliki nilai kemiripan dengan data uji sebagai opsi rekomendasi. Hal ini memungkinkan tidak jarang hasil rekomendasi yang ditawarkan kurang sesuai dengan kebutuhan data uji. Pada penelitian ini, peneliti akan melakukan analisa mengenai perbandingan beberapa teknik perhitungan similiaritas profil mahasiswa, dan menggabungkan *collaborative filtering* dengan *K-Means Clustering* (pendekatan pengelompokan atau klusterisasi data menjadi beberapa kelas berdasarkan kemiripan data) untuk menghasilkan nilai akurasi rekomendasi yang lebih optimal dengan meminimalisir keluarnya rekomendasi mata kuliah yang kurang sesuai dengan profil data uji.

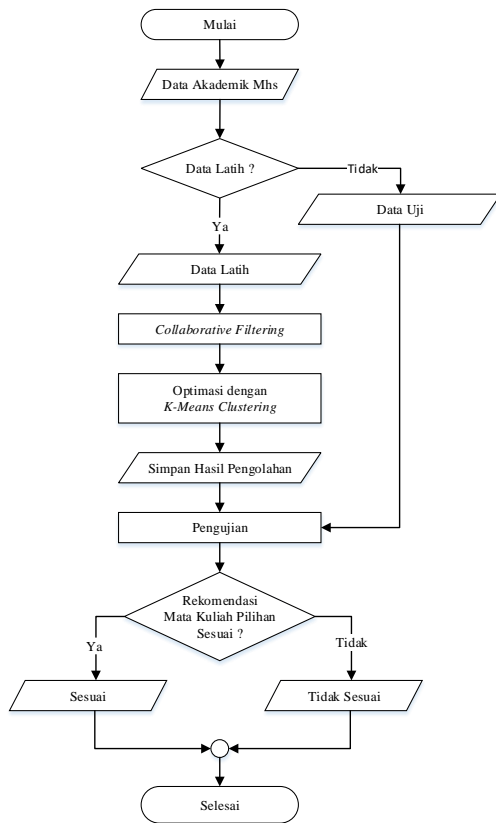
2. METODE PENELITIAN

Pada tahap ini langkah-langkah penelitian yang dilakukan sesuai dengan alur penelitian pada gambar 1 adalah sebagai berikut :

A. Teknik Pengambilan Data

Proses pengambilan data dalam penelitian ini adalah dengan mengeliminasi data yang memiliki atribut *redundant* seperti pada kasus mahasiswa yang mengulang mata kuliah, mengeliminasi data yang atributnya tidak lengkap seperti pada kasus mahasiswa yang tidak melanjutkan studi. Untuk penelitian ini, data yang digunakan adalah data

akademik mahasiswa pada dua tahun angkatan pada satu program studi di salah satu universitas yang memiliki tahun mata kuliah yang sama. Hal ini untuk mengasumsikan kurikulum yang digunakan adalah kurikulum yang sama, sehingga ada relasi kesesuaian kode mata kuliah pada data set dan data uji. Dari persiapan tersebut, diperoleh 118 data latih yaitu data akademik untuk mahasiswa angkatan 1 dan 92 data uji yaitu data akademik untuk mahasiswa angkatan 2. Atribut yang digunakan untuk pengujian antara lain kode mata kuliah, nilai mata kuliah yang diasumsikan sebagai bobot, semester, dan konsentrasi prodi. Data yang akan diuji akurasi adalah data prediksi konsentrasi prodi dan akurasi berapa persen mata kuliah yang dianjurkan oleh sistem yang pada akhirnya diambil pula oleh data mahasiswa yang diuji.



Gambar 1. Diagram Alur Penelitian

B. Collaborative Filtering

Proses implementasi pendekatan *collaborative filtering* akan dilakukan dengan empat metode, yaitu *jaccard*, *euclidean distance*, *cosine similarity*, dan *pearson correlation*. Adapun penjabarannya sebagai berikut :

1) *Jaccard*

Perhitungan *similarity* pada metode *jaccard* ini cukup sederhana. Hal ini dikarenakan data latih dan data uji hanya perlu dibandingkan apakah mereka berdua memiliki atribut yang sama atau tidak [9]. Dalam hal pengimplementasiannya pada program

perekomendasi mata kuliah ini, tiap data set dan data uji akan dibandingkan apakah kedua data mahasiswa memiliki kesamaan pengambilan mata kuliah sebelumnya. Untuk seluruh agregat kode mata kuliah yang ada, mata kuliah yang pernah diambil oleh data set dan data uji akan ditandai dengan label 1, sedangkan untuk kode mata kuliah yang belum diambil akan diberi label 0. Dengan demikian nilai *similarity* dari kedua data tersebut akan diperoleh dan diurutkan. Nilai yang paling mendekati 1 maka diasumsikan sebagai data set yang paling memiliki kemiripan dengan profil data uji, sehingga mata mata kuliah semester selanjutnya akan menjadi bahan rekomendasi untuk diambil oleh data uji. Pada persamaan (1) akan dijabarkan persamaan untuk memperoleh nilai *jaccard similarity* [9].

$$J_{sim}(u, u') = \frac{|u \cap u'|}{|u \cup u'|}$$

Keterangan :

- $J_{sim}(u, u')$ = Nilai *similarity*
- $u \cap u'$ = Jumlah irisan vektor data uji dan data latih
- $u \cup u'$ = Jumlah gabungan vektor data uji dan data latih

2) *Euclidean Distance*

Perhitungan *similarity* pada metode *euclidean distance* ini menggunakan jarak *node* sebagai parameter kemiripan [10]. Semakin kecil jarak antar 2 *node*, maka kedua *node* tersebut dapat diasumsikan memiliki tingkat kemiripan yang tinggi. Dalam hal pengimplementasiannya pada program rekomendasi mata kuliah ini, tiap data set dan data uji akan dibandingkan antara bobot-bobot mata kuliah sebelumnya. Bobot mata kuliah diambil melalui nilai yang diperoleh dari mata kuliah yang telah diambil dan dikonversikan dari nilai huruf ke nilai angka, skala 8-1 untuk nilai A-E. Konsep perhitungan *euclidean distance* menggunakan rumus *pythagoras*, dimana jaraknya dapat diperoleh dengan menghitung akar kuadrat dari sigma selisih jarak *node* pangkat dua. Nilai yang paling mendekati 0 maka diasumsikan sebagai data set yang paling memiliki kemiripan dengan profil data uji, sehingga mata mata kuliah semester selanjutnya akan menjadi bahan rekomendasi untuk diambil oleh data uji. Pada persamaan (2) akan dijabarkan persamaan untuk memperoleh nilai *euclidean distance* [10].

$$Euc(u, u') = \sqrt{\sum_{i=0}^n (u - u')^2}$$

Keterangan :

- $Euc(u, u')$ = Nilai *euclidean distance*
- $u - u'$ = Nilai absolut selisih bobot vektor data uji dan data latih

3) *Cosine Similarity*

Sama halnya dengan perhitungan *euclidean distance*, *cosine similarity* ini juga menggunakan jarak *node* sebagai parameter kemiripan. Yang membedakan dari *euclidean distance* adalah, *cosine similarity* akan menghitung derajat kemiripan dari

vektor-vektor yang dimiliki setiap data [11]. Nilai perhitungan derajat kemiripannya dapat bervariasi mulai dari 0 hingga 1. Semakin nilai *cosine similarity*-nya mendekati angka 1, maka kedua *node* tersebut dapat diasumsikan memiliki tingkat kemiripan yang tinggi. Sama seperti konsep perhitungan kosinus, dimana Cos 0 derajat adalah 1, maka diperlukan selisih jarak mendekati 0 derajat untuk menghasilkan nilai mendekati 1 (kemiripan tertinggi). Dalam hal pengimplementasiannya pada program perkomendasi mata kuliah ini, tiap data set dan data uji akan dibandingkan antara bobot-bobot mata kuliah sebelumnya. Bobot mata kuliah diambil melalui nilai yang diperoleh dari mata kuliah yang telah diambil dan dikonversikan dari nilai huruf ke nilai angka, skala 8-1 untuk nilai A-E. Perhitungan *cosine similarity* menggunakan rumus dimana jaraknya dapat diperoleh dengan membagi jumlah *dot product* dari setiap vektor dengan *dot product* dari *euclidean distance* masing-masing *node*. Nilai yang paling mendekati 1 maka diasumsikan sebagai data set yang paling memiliki kemiripan dengan profil data uji, sehingga mata kuliah semester selanjutnya akan menjadi bahan rekomendasi untuk diambil oleh data uji. Pada persamaan (3) akan dijabarkan persamaan untuk memperoleh nilai *cosine similarity* [11].

$$C_{sim}(u, u') = \frac{\sum_{i=0}^n u_i w_i}{\sqrt{\sum_{i=0}^n u_i^2} \sqrt{\sum_{i=0}^n w_i^2}}$$

Keterangan :

$$\begin{aligned} C_{sim}(u, u') &= \text{Nilai } cosine \text{ similarity} \\ u - u' &= \text{Nilai absolut selisih bobot} \\ &\quad \text{vektor data uji dan data latihan} \\ n &= \text{Jumlah data} \end{aligned}$$

4) Pearson Correlation

Berbeda dengan pendekatan sebelumnya, *pearson correlation* digunakan untuk menghitung ukuran korelasi linier antara dua set data (*node*) [12]. Pendekatan ini akan menghitung rasio antara kovarian dari dua vektor dan *dot product* dari simpangan bakunya, atau dengan kata lain perhitungan ini akan mengukur kovarian yang dinormalisasi. Nilai perhitungan derajat kemiripannya dapat bervariasi mulai dari -1 hingga 1. Semakin nilai *pearson correlation*-nya mendekati angka 1, maka kedua *node* tersebut dapat diasumsikan memiliki tingkat kemiripan yang tinggi. Nilai 0 menandakan kedua *node* tidak memiliki korelasi sama sekali, sedangkan nilai -1 menandakan kedua *node* berkorelasi namun sangat bertolak belakang. Dalam hal pengimplementasiannya pada program perkomendasi mata kuliah ini, tiap data set dan data uji akan dibandingkan antara bobot-bobot mata kuliah sebelumnya. Bobot mata kuliah diambil melalui nilai yang diperoleh dari mata kuliah yang telah diambil dan dikonversikan dari nilai huruf ke nilai angka, skala 8-1 untuk nilai A-E. Nilai yang paling mendekati 1 maka diasumsikan sebagai data set yang paling memiliki kemiripan dengan profil data uji, sehingga mata kuliah semester

selanjutnya akan menjadi bahan rekomendasi untuk diambil oleh data uji. Pada persamaan (4) akan dijabarkan persamaan untuk memperoleh nilai *pearson correlation* [12].

$$P_{sim}(u, u') = \frac{n \sum_{i=0}^n u_i w_i - \sum_{i=0}^n u_i \sum_{i=0}^n w_i}{\sqrt{n \sum_{i=0}^n u_i^2 - (\sum_{i=0}^n u_i)^2} \sqrt{n \sum_{i=0}^n w_i^2 - (\sum_{i=0}^n w_i)^2}}$$

Keterangan :

$$\begin{aligned} P_{sim}(u, u') &= \text{Nilai } pearson \text{ correlation} \\ u - u' &= \text{Nilai absolut selisih bobot} \\ &\quad \text{vektor data uji dan data latihan} \\ n &= \text{Jumlah data} \end{aligned}$$

C. K-Means Clustering

K-Means clustering merupakan salah satu proses klasterisasi dimana pendekatan ini akan mengkategorikan sebuah data dalam beberapa kelas untuk memudahkan dalam mengambil suatu keputusan. Pembagian kelas pada suatu data tersebut dilakukan dengan cara melakukan iterasi data dan mencari jarak tiap datanya menggunakan *euclidean distance*. Pada saat proses iterasi perbandingan data, setiap data dapat berubah posisi atau kelas sesuai dengan perubahan nilai *centroid* dari tiap iterasi. Pada penelitian sebelumnya, K-Means dapat melakukan proses kasterisasi suatu objek dengan baik [13] [14]. Pada penelitian ini, beberapa kali mata kuliah hasil rekomendasi dianggap terlalu banyak. Hal ini tentu saja mempengaruhi nilai akurasi perbandingan mata kuliah yang direkomendasikan dengan mata kuliah yang akhirnya diambil. Oleh karena itu, *K-Means clustering* akan membantu mengeliminasi mata kuliah mana saja yang sebaiknya direkomendasikan dan tidak, untuk meningkatkan nilai akurasi.

Langkah yang dilakukan dalam penelitian ini adalah melakukan *query* pada data untuk melakukan agregasi jumlah pengambil mata kuliah tiap konsentrasi prodi pada semester yang ingin direkomendasikan. Dari hasil *query* tersebut maka akan diketahui jumlah agregat mahasiswa yang memiliki atribut mata kuliah tertentu pada semester terkait. Setelah itu *K-Means clustering* akan melakukan proses klasterisasi untuk membagi agregat data tersebut kedalam dua kelas yang menandai sebuah mata kuliah direkomendasikan atau tidak direkomendasikan. Berikut adalah langkah-langkah implementasi *K-Means clustering* pada penelitian ini :

1. Tentukan jumlah kelas yang diinginkan. Dalam hal ini, kelas akan dibagi menjadi 2 untuk menentukan label direkomendasikan dan tidak direkomendasikan.
2. Karena ada 2 kelas, pilih 2 data sejumlah kelas yang diinginkan yang akan digunakan sebagai *centroid*.
3. Hitung jarak semua data dengan 2 *centroid* dengan menggunakan *euclidean distance* [13] dengan notasi sebagai berikut :

$$D_{(i,j)} = \sqrt{(X_{1i} - X_{1j})^2 + \dots + (X_{ki} - X_{kj})^2} \dots\dots\dots(5)$$

Keterangan :

- $D_{(i,j)}$ = Euclidean distance / Jarak data i ke pusat *centroid* kelas j
- X_{ki} = Data ke- i pada atribut data k
- X_{kj} = *Centroid* ke- j pada atribut data k

4. Hasil perhitungan jarak akan dibandingkan dengan tiap *centroid*, data yang memiliki jarak paling dekat, akan dikelompokkan menjadi satu kelas.
5. Tentukan *centroid-centroid* baru yang diperoleh dari rata-rata (*means*) data tiap kelas yang baru saja terbentuk.
6. Ulangi proses hitung jarak pada pengelompokan kelas baru hingga nilai *centroid* tidak berubah lagi dan kelas akhir dapat ditentukan.
7. Setelah selesai menentukan kelas dari setiap agregat konsentrasi, seluruh hasil klusterisasi dimasukkan kedalam tabel keputusan untuk menjadi dasar keputusan rekomendasi.

D. Analisa Hasil

Tahap akhir dari penelitian ini adalah melakukan evaluasi terhadap kinerja sistem rekomendasi. Evaluasi sistem rekomendasi ini menggunakan *Moving Average* (MA) untuk menghitung akurasi pengklasifikasian konsentrasi perkuliahan dan akurasi pengusulan mata kuliah [15]. Selain itu penelitian ini juga akan membandingkan *Mean Absolute Error* (MAE) untuk evaluasi keakuratan untuk menghitung kesalahan mutlak dari rekomendasi mata kuliah yang muncul.

Adapun perhitungan *Moving Average* [15] dapat dinotasikan sebagai berikut :

$$MA = \frac{A_1 + A_2 + A_3 + \dots + A_n}{n} \dots\dots\dots(6)$$

Keterangan :

- MA = Nilai *moving average*
- A_n = Nilai ke- n
- n = Jumlah data yang diproses

Sedangkan perhitungan *Mean Absolute Error* [15] dapat dinotasikan sebagai berikut :

$$MAE = \frac{\sum_{i=0}^n |A_i - P_i|}{n} \dots\dots\dots(7)$$

Keterangan :

- MAE = Nilai *mean absolute error*
- A_i = Nilai aktual
- P_i = Nilai Prediksi
- n = Jumlah data yang diproses

3. HASIL DAN PEMBAHASAN

Pada penelitian ini, telah diimplementasikan suatu sistem rekomendasi mata kuliah dengan empat metode pendekatan *collaborative filtering*. Dari hasil yang diperoleh masing-masing metode pendekatan, akan difokuskan pada nilai *moving average* dari kesesuaian sistem memunculkan profil mahasiswa

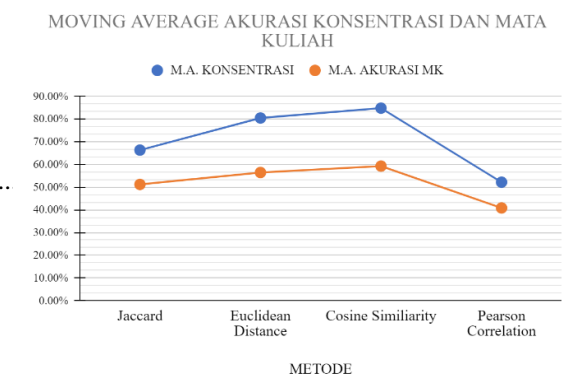
yang memiliki konsentrasi yang sesuai dengan data uji, dan juga menghitung *moving average* dari akurasi mata kuliah yang direkomendasikan. Dari implementasi awal hasil perhitungan dapat dilihat pada tabel 1.

Nilai akurasi *moving average* pada akurasi konsentrasi didapatkan dari perbandingan antara jumlah konsentrasi yang sesuai dengan jumlah keseluruhan data uji. Sedangkan nilai akurasi *moving average* pada akurasi rekomendasi mata kuliah didapatkan dari perbandingan jumlah mata kuliah yang sesuai dan jumlah mata kuliah yang disarankan secara keseluruhan, setelah itu dilakukan perhitungan rata-rata.

Tabel 1. Nilai *Moving Average* Implementasi Awal *Collaborative Filtering*

Metode	MA. Konsentrasi	MA. Akurasi MK
<i>Jaccard</i>	66.30%	53.66%
<i>Euclidean Distance</i>	80.43%	56.44%
<i>Cosine Similarity</i>	84.78%	59.25%
<i>Pearson Correlation</i>	52.17%	43.71%

Dari grafik diatas dapat diketahui bahwa pendekatan yang menghasilkan nilai akurasi tertinggi baik untuk kesesuaian konsentrasi maupun mata kuliah yang berhasil direkomendasikan jatuh pada metode *cosine similarity* dengan akurasi konsentrasi 84.78% dan akurasi rekomendasi mata kuliah 59.25%, disusul oleh metode *euclidean distance* dengan akurasi konsentrasi 80.43% dan akurasi rekomendasi mata kuliah 56.44%, kemudian *jaccard* dengan akurasi konsentrasi 66.30% dan akurasi rekomendasi mata kuliah 53.66% dan nilai paling rendah jatuh pada metode *pearson correlation* dengan akurasi konsentrasi 52.17% dan akurasi rekomendasi mata kuliah 43.71%..... (6)



Gambar 2. Perbandingan Nilai Akurasi *Moving Average* Konsentrasi dan Akurasi Mata Kuliah

Pada keempat pendekatan *collaborative filtering* tersebut, terdapat fenomena pada metode *jaccard*, dimana banyak data hasil perhitungan yang memiliki nilai *similarity* terbaik lebih dari satu. Hal ini menyebabkan sistem tidak dapat menentukan data

rekomendasi yang cukup akurat dikarenakan munculnya banyak profil dari data set yang memiliki *similarity* namun memiliki profil yang berbeda. Hasil perhitungan terkoreksi manual menampilkan nilai akurasi yang tertinggi pada 83.69% pada konsentrasi dan 61.06% pada akurasi rekomendasi mata kuliah. Hal ini dimungkinkan karena *jaccard* hanya akan mengenali nilai *true* dan *false* dalam perbandingan nilai vektor, dalam hal ini seperti contohnya mahasiswa uji dengan mahasiswa latih sama-sama telah menempuh mata kuliah X, namun mahasiswa uji mendapatkan nilai A sedangkan mahasiswa latih mendapatkan nilai D. Dalam kasus ini, *jaccard* akan mengenali mereka berdua memiliki kemiripan tinggi karena sama-sama pernah menempuh mata kuliah X walaupun pada kenyataannya profil keduanya sangat berbeda.

Untuk pendekatan *collaborative filtering* menggunakan metode *cosine similarity* mendapatkan nilai akurasi yang tertinggi diantara keempat metode lainnya dan disusul oleh metode *euclidean distance* dengan selisih akurasi yang cukup rendah. *Cosine similarity* berhasil mendapatkan nilai akurasi yang paling tinggi dikarenakan perhitungan yang dilakukan akan membandingkan sudut derajat kesamaan dari vektor-vektor yang ada, bukan besar jarak antara vektor-vektor. Sedangkan perhitungan *euclidean distance* akan menghitung jarak tiap vektor untuk menentukan mana jarak terdekat dari data uji dan data latih. Dalam penelitian ini dapat diasumsikan dengan skenario seperti berikut :

- Nilai Mahasiswa Uji :
 - MK-A = A
 - MK-B = A
 - MK-C = A
- Nilai Mahasiswa Latih-1 :
 - MK-A = C
 - MK-B = B
 - MK-C = A
- Nilai Mahasiswa Latih-2 :
 - MK-A = A
 - MK-B = A
 - MK-D = A

Pada skenario diatas, pendekatan *cosine similarity* akan menghasilkan mahasiswa latih-1 sebagai nilai yang memiliki *similarity* yang tinggi. Hal ini karena meskipun mahasiswa uji dengan mahasiswa latih 1 memiliki bobot nilai berbeda, namun mereka sama-sama pernah memilih mata kuliah yang sama, yaitu MK-A, MK-B, dan MK-C. Untuk perhitungan *euclidean distance*, maka nilai yang memiliki jarak paling kecil yang dianggap memiliki derajat kemiripan tertinggi dengan mahasiswa uji adalah mahasiswa latih-2. Hal ini dikarenakan meskipun tidak semua mata kuliah yang pernah diambil mutlak sama, namun keduanya sama-sama memiliki semua nilai mata kuliah A.

Untuk pendekatan yang memiliki nilai akurasi terendah adalah pada penggunaan metode *pearson*

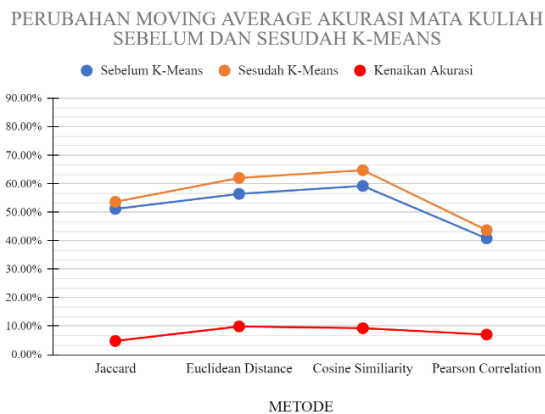
correlation. Pada metode ini, secara umum vektor-vektor akan dibandingkan dan dicari skala korelasinya. Perhitungan metode ini secara garis besar akan mengasumsikan banyaknya korelasi data akan mendukung hasil prediksi untuk kejadian yang selanjutnya. Pada penelitian ini, skenario yang memungkinkan adalah dimana data mahasiswa uji memiliki nilai antara mata kuliah MK-A = A, MK-B = A, dan MK-C = A, sedangkan mahasiswa latih memiliki nilai MK-A = E, MK-B = E, dan MK-C = E. Untuk perbedaan nilai ekstrem tersebut, nilai *pearson correlation* akan menghasilkan nilai negatif, dimana kedua vektor tersebut berkorelasi dengan ditandai dengan pengambilan mata kuliah yang sama, namun memiliki nilai yang bertolak belakang. Selain itu, terdapat beberapa kondisi dimana penggunaan *pearson correlation* tidak optimal, antara lain dikarenakan data yang tidak linear, sedikitnya *outlier*, data memiliki beberapa *subgroup*, dan data memiliki *heteroscedasticity* atau situasi di mana satu variabel memiliki variabilitas yang tidak sama di seluruh rentang nilai variabel kedua [16].

Setelah pengaplikasian keempat pendekatan *collaborative filtering* telah diterapkan, selanjutnya adalah penerapan *K-Means clustering* untuk melakukan optimalisasi akurasi mata kuliah yang akan direkomendasikan. Dari hasil perhitungan *moving average* didapatkan perubahan nilai akurasi yang dapat dilihat pada tabel 2.

Dari penelitian ini, pendekatan *K-Means clustering* mampu meningkatkan akurasi perekomendasi mata kuliah dengan metode *jaccard* sebesar 4.83%, metode *euclidean distance* sebesar 9.92%, *cosine similarity* sebesar 9.30% dan *pearson correlation* sebesar 7.05%. Dengan mengaplikasikan pendekatan *K-Means clustering*, sistem mampu mereduksi pemunculan rekomendasi mata kuliah yang telah di klasterisasi oleh *K-Means* sehingga nilai akurasi *moving average* dari mata kuliah yang muncul dapat lebih dioptimalkan. Walaupun pada kasus-kasus tertentu terdapat mata kuliah yang muncul dalam rekomendasi dan juga diambil oleh mahasiswa uji, namun mata kuliah tersebut harus tidak direkomendasikan oleh sistem karena klasterisasi *K-Means* telah memfilter mata kuliah tersebut untuk tidak direkomendasikan.

Tabel 2. Nilai Kenaikan Akurasi Pendekatan *Collaborative Filtering* dengan *K-Means Clustering*

Metode	Moving Average Akurasi MK		Kenaikan Akurasi
	Sebelum <i>K-Means</i>	Sesudah <i>K-Means</i>	
<i>Jaccard</i>	51.18%	53.66%	4.83%
<i>Euclidean Distance</i>	56.44%	62.04%	9.92%
<i>Cosine Similarity</i>	59.25%	64.76%	9.30%
<i>Pearson Correlation</i>	40.83%	43.71%	7.05%



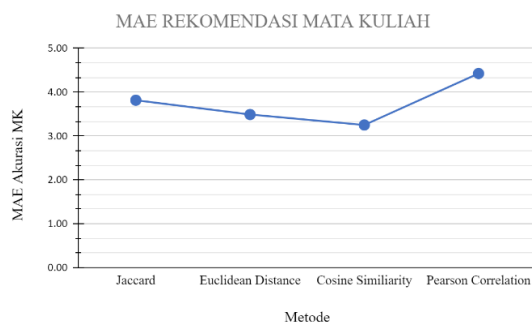
Gambar 3. Perubahan Nilai Akurasi Mata Kuliah Sebelum dan Sesudah *K-Means Clustering*

Untuk selanjutnya, perhitungan *Mean Absolute Error* dari keempat metode yang telah teroptimalisasi dapat dilihat dari tabel 3 berikut :

Tabel 3. Nilai *Mean Absolute Error* Akurasi Mata Kuliah

Metode	MAE Akurasi MK
<i>Jaccard</i>	3.82
<i>Euclidean Distance</i>	3.49
<i>Cosine Similarity</i>	3.25
<i>Pearson Correlation</i>	4.42

Dari penelitian ini, dihitung juga nilai *Mean Absolute Error* dari akurasi perekomendasi mata kuliah. Nilai *Mean Absolute Error* digunakan untuk menghitung nilai rata-rata selisih mutlak nilai aktual dan nilai prediksi. Semakin kecil nilai yang diperoleh, maka semakin optimum pendekatan metode yang dilakukan. Dalam penelitian ini diperoleh nilai MAE untuk metode *jaccard* teroptimisasi sebesar 3.82, *euclidean distance* teroptimisasi sebesar 3.49, *cosine similarity* teroptimisasi sebesar 3.25, dan *pearson correlation* teroptimisasi sebesar 4.42. Dengan demikian dapat disimpulkan bahwa pendekatan metode terbaik adalah pada metode *cosine similarity* teroptimisasi dan pendekatan metode yang paling tidak optimal adalah pada metode *pearson correlation* teroptimisasi.



Gambar 4. Nilai *Mean Absolute Error* Rekomendasi Mata Kuliah Setelah optimasi *K-Means Clustering*

4. KESIMPULAN

Dalam penelitian ini, sistem rekomendasi berbasis *collaborative filtering* dibangun dengan empat metode, yaitu *jaccard*, *euclidean distance*, *cosine similarity*, dan *pearson correlation*. Dari pengimplementasian keempat metode tersebut, sistem mampu melakukan prediksi terkait konsentrasi prodi mahasiswa dan juga rekomendasi mata kuliah yang dapat diambil pada semester selanjutnya. Dari implementasi awal, didapatkan hasil bahwa metode yang paling optimal hingga yang kurang optimal secara berurutan adalah metode, *cosine similarity*, *euclidean distance*, *jaccard*, lalu *pearson correlation*.

Proses optimalisasi hasil rekomendasi mata kuliah pada penelitian ini dilakukan dengan melakukan seleksi rekomendasi berdasarkan tabel keputusan yang diolah menggunakan metode *K-Means clustering*. *K-means clustering* mampu melakukan klusterisasi dari seluruh mata kuliah yang muncul dalam rekomendasi dan memberikan label untuk mana mata kuliah yang direkomendasikan dan tidak direkomendasikan. Pada pengimplementasiannya, tabel keputusan *K-Means clustering* mampu meningkatkan nilai akurasi *moving average* pada setiap metode pendekatan *collaborative filtering* dengan persentase peningkatan akurasi secara berurutan dari yang paling optimal hingga kurang optimal yaitu *euclidean distance* sebesar 9.92%, *cosine similarity* sebesar 9.30%, *pearson correlation* sebesar 7.05%, dan *jaccard* sebesar 4.83%.

Secara keseluruhan, sistem rekomendasi mata kuliah ini mampu melakukan prediksi konsentrasi mahasiswa hingga 84.78% dan akurasi *moving average* pada prediksi rekomendasi mata kuliah hingga 64.76% menggunakan data latih yang ditentukan. Dengan nilai *mean absolute error* yang dihasilkan pada penelitian ini, pendekatan *collaborative filtering* yang paling optimal untuk perekomendasi mata kuliah jatuh pada metode *cosine similarity*.

Meskipun demikian, penelitian ini masih memiliki banyak kekurangan. Selain nilai akurasi rekomendasi mata kuliah yang cukup rendah, penggunaan data latih dengan nilai skalabilitas yang baik memungkinkan menghasilkan nilai prediksi yang lebih baik meskipun akan menambah *cost* untuk melakukan penyetaraan kurikulum mata kuliah. Penggunaan kuesioner kepuasan perkuliahan bisa menjadi alternatif untuk memperoleh preferensi keminatan mahasiswa dibandingkan dengan menggunakan data akademik yang lebih mengacu kepada keberhasilan mahasiswa.

5. DAFTAR PUSTAKA

- [1] G. Engin, B. Aksoyer, D. Bozanli, U. Hanay, D. Maden and G. Ertek. 2014. "Rule-based Expert System for Supporting University

- Students," in *2nd International Conference on Information Technology and Quantitive Management*, Istanbul, Turkey.
- [2] H. Bydovská and L. Popelínský. 2013. "Predicting Student Performance in Higher Education," in *24th International Workshop on Database and Expert Systems Applications*, Los Alamitos, CA, USA.
- [3] S. B. Sailesh, K. J. Lu and M. A. Aali. 2016. "Profiling students on their course-taking patterns in Higher Educational Institutions (HEIs)," in *International Conference on Information Science (ICIS)*, Kochi, India.
- [4] P. Mudjiraharjo, A. Fauzi and H. Tolle. 2018. "The Recommendation System of Thesis Topics Selection Based on Fuzzy-AHP and Fuzzy-ANP (Case Study: D-IV Nursing Program of Health Polytechnic, Department of Health, Malang)," in *The 9th International Electrical Power, Electronics, Communications, Control, and Informatics Seminar (EECCIS)*, Malang.
- [5] J. S. Yoo, Y.-S. Woo and S. J. Park. 2017. "Mining Course Trajectories of Successful and Failure Students: A Case Study," in *2017 IEEE International Conference on Big Knowledge (ICBK)*, Hefei, China.
- [6] S. Srivastava, S. Karigar, R. Khanna and R. Agarwal. 2018. "Educational Data Mining: Classifier Comparison for the Course Selection Process," in *International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*.
- [7] T. Cunha, C. Soares and A. de Carvalho. 2018. "Metalearning and Recommender Systems: A literature review and empirical study on the algorithm selection problem for Collaborative Filtering," *Information Sciences*, pp. 128-144.
- [8] A. Cakmak. 2017. "Predicting Student Success in Courses via Collaborative Filtering," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 5, no. 1, pp. 10-17.
- [9] A. Rana and K. Deeba. 2019. "Online Book Recommendation System using Collaborative Filtering (With Jaccard Similarity)," in *International Conference on Physics and Photonics Processes in Nano Sciences*, Eluru, India.
- [10] D. Adeniyi, Z. Wei and Y. Yongquan. 2016. "Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method," *Applied Computing and Informatics*, vol. 12, no. 1, pp. 90-108.
- [11] A. R. Lahitani. 2022. "Automated Essay Scoring menggunakan Cosine Similarity pada Penilaian Esai Multi Soal," *Jurnal Kajian Ilmiah*, vol. 22, no. 2, pp. 107-118.
- [12] B. Prasetyo, H. Haryanto, S. Astuti, E. Z. Astuti and Y. Rahayu. 2019. "Implementasi Metode Item-Based Collaborative Filtering dalam Pemberian Rekomendasi Calon Pembeli Aksesoris Smartphone," *Jurnal Eksplorasi Informatika*, vol. 9, no. 1, pp. 17-27.
- [13] R. J. Hablum, A. Khairan and Rosihan. 2019. "Clustering Hasil Tangkapan Ikan di Pelabuhan Perikanan Nusantara (PNN) Ternate Menggunakan Algoritma K-Means," *Jurnal Informatika dan Komputer*, vol. 02, no. 1, pp. 26-33.
- [14] R. Adhitama, A. Burhanuddin and R. Ananda. 2020. "Penentuan Jumlah Cluster Ideal SMK di Jawa Tengah dengan Metode X-Means Clustering dan K-Means Clustering," *Jurnal Informatika dan Komputer*, vol. 3, no. 1, pp. 1-5.
- [15] H. Prapcoyo. 2018. "Peramalan Jumlah Mahasiswa Menggunakan Moving Average," *Telematika*, vol. 15, no. 01, pp. 66-75.
- [16] R. Aggarwal and P. Ranganathan. 2016. "Common pitfalls in statistical analysis: The use of correlation techniques," *Perspectives in clinical research Vol. 7,4*, pp. 187-190.