

Sistem Rekomendasi Mata Kuliah Pilihan Menggunakan Metode User Based Collaborative Filtering Berbasis Algoritma Adjusted Cosine Similarity

Glenn Ferio, Rolly Intan, Silvia Rostianingsih

Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Kristen Petra

Jl. Siwalankerto 121 – 131 Surabaya 60236

Telp. (031) – 2983455, Fax. (031) – 8417658

E-Mail: glennferiof@gmail.com, rintan@petra.ac.id, silvia@petra.ac.id

ABSTRAK

Sistem rekomendasi merupakan sistem yang memberikan rekomendasi pada suatu item yang dapat digunakan untuk membantu *user* dalam mengambil keputusan. Dalam perkuliahan, mahasiswa diwajibkan untuk mengambil keputusan akan mata kuliah pilihan yang harus diambil selama menjalani perkuliahan. Akan tetapi, tidak jarang banyak sekali mahasiswa yang kebingungan dalam menentukan mata kuliah pilihan yang harus diambil. Dengan demikian diperlukan suatu sistem yang dapat memberikan rekomendasi mata kuliah pilihan sehingga dapat membantu mahasiswa dalam pengambilan keputusan akan mata kuliah pilihan yang akan diambil.

Salah satu metode yang sangat terkenal untuk dipergunakan pada sistem rekomendasi adalah *Collaborative Filtering*, dimana dalam metode ini dilakukan kalkulasi nilai kemiripan dari *user* yang akan diprediksi dengan *user-user* lainnya. Metode ini menggunakan Algoritma *Adjusted Cosine Similarity* dan *K-Nearest Neighbors* dalam menentukan *user-user* mana yang memiliki kemiripan terbaik dengan *user* sehingga dari data *user-user* inilah dapat dilakukan prediksi pada mata kuliah pilihan yang belum diambil oleh *user*.

Hasil dari pengujian yang dilakukan, ditemukan algoritma *Adjusted Cosine Similarity* memiliki performa yang lebih dari *Cosine Similarity* dalam memberikan rekomendasi mata kuliah pilihan. Pada pengujian pada *K-Nearest Neighbors*, akurasi terbaik berada pada nilai variabel $K=16$ pada *K-Nearest Neighbors* dengan akurasi sebesar 89.31422 %.

Kata Kunci: *Recommender System, Collaborative Filtering, Adjusted Cosine Similarity, K-Nearest Neighbors.*

ABSTRACT

The recommendation system is a system that provides recommendations on an item that can be used to assist users in making decisions. In lectures, students as users are required to make decisions on elective courses that must be taken during the course. However, it is not uncommon for many students to be confused in deciding which elective courses to take. Thus a system is needed that can provide recommendations for elective courses so that it can help students in making decisions on choosing the best elective courses.

One very well-known method to use on recommendation systems is Collaborative Filtering, wherein this method will calculate the similarities from user (that will be predicted) with other users. This method uses the Adjusted Cosine Similarity and K-Nearest Neighbors Algorithm in determining which users have the best similarity with user (that will be predicted) so that from this user,

predictions can be made on the elective courses that have not been taken by the user.

The result shows that the Adjusted Cosine Similarity algorithm works more better than Cosine Similarity in providing recommendations for elective courses. In testing on K-Nearest Neighbors, the best accuracy is in the variable value $K = 16$ on K-Nearest Neighbors with an accuracy of 89.31422%.

Keywords: *Recommender System, Collaborative Filtering, Adjusted Cosine Similarity, K-Nearest Neighbors*

1. PENDAHULUAN

Program Studi (Prodi) Teknik Informatika merupakan salah satu program studi di UK Petra dimana dalam prodi ini, mahasiswa akan diajarkan mengenai berbagai kompetensi yang diperlukan dalam bidang IT (*Information and Technology*) sehingga dapat mengaplikasikannya untuk menyelesaikan berbagai persoalan pada bidang IT. Untuk menempuh kelulusan pada prodi ini, mahasiswa dituntut untuk dapat menyelesaikan 144 sks (Satuan Kredit Semester) dalam kurun waktu 3.5 – 5 tahun. Selama menempuh pendidikan di Prodi Teknik Informatika, mahasiswa akan diminta untuk menyelesaikan berbagai mata kuliah wajib dan memiliki kebebasan untuk memilih mata kuliah pilihan yang disediakan oleh Prodi Teknik Informatika.

Suatu sistem rekomendasi dibuat untuk memberikan alternatif pertimbangan yang berguna untuk mendukung keputusan mahasiswa dalam menentukan mata kuliah yang ingin diambil. Hal ini mendasari diperlukannya suatu sistem rekomendasi dengan akurasi terbaik sesuai dengan karakteristik data dari *database* mahasiswa Prodi Teknik Informatika. Pada penelitian yang dilakukan oleh berbagai sumber akurasi dari berbagai penelitian bervariasi karena bergantung pada faktor yang paling sesuai dengan karakteristik data yang ada. Hal ini mendasari permasalahan yang dalam penentuan suatu akurasi yang terbaik yang bisa dihasilkan dari data yang tersedia.

Penelitian akan dilakukan terhadap pengaruh faktor jumlah *Top-N similarity* dan konversi pada nilai mata kuliah pilihan dengan perbedaan kurikulum dalam mempengaruhi akurasi dari *Collaborative Filtering* sehingga dapat memberikan rekomendasi 10 mata kuliah pilihan berdasarkan pada nilai prediksi mata kuliah pilihan yang terbaik. Dalam hal ini perhitungan akurasi dari nilai prediksi dilakukan dengan menggunakan MAE (*Mean Average Error*) dan penelitian dilakukan dengan mempergunakan algoritma *Adjusted Cosine Similarity*. Penggambaran dari hasil testing akan dilakukan dalam bentuk grafik dan tabel perbandingan akurasi dari metode *Collaborative Filtering* dengan algoritma *Adjusted Cosine Similarity*.

2. DASAR TEORI

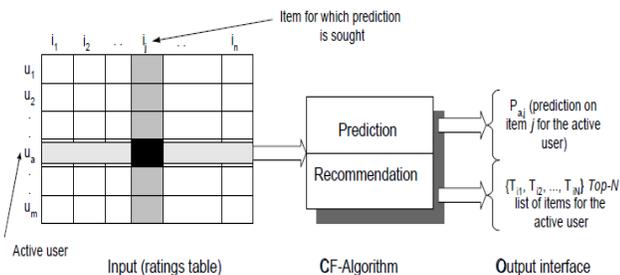
2.1 Collaborative Filtering Sebagai Sistem Rekomendasi

Sistem rekomendasi merupakan suatu sistem yang dapat memberikan informasi dan rekomendasi yang membantu *user* dalam membuat keputusan berdasarkan data yang telah ada sebelumnya.

Salah satu metode yang dipergunakan dalam pembuatan suatu sistem rekomendasi yang terkenal hingga saat ini adalah *Collaborative Filtering*. Pada implementasi *Collaborative Filtering* dalam sistem rekomendasi mata kuliah pilihan, data yang dibandingkan akan disimpan dalam sebuah matriks dua dimensi dimana dimensi dari y atau baris akan menyatakan mahasiswa dan dimensi x atau kolom akan menyatakan mata kuliah.

2.2 Proses Collaborative Filtering

Ide dasar dari menghitung *similarity* antara mahasiswa U_1 dengan U_m adalah dengan membandingkan data semua nilai mata kuliah yang dari mahasiswa U_1 dengan U_m untuk menghasilkan nilai prediksi pada mata kuliah yang belum pernah diambil oleh mahasiswa U_m . Penerapan metode ini dilakukan dengan mencari semua *user* aktif yang memiliki nilai *Top-K similarity* paling tinggi dengan mahasiswa U_m . Setelah ditemukan *Top-K* mahasiswa yang memiliki nilai *similarity* tertinggi dengan mahasiswa U_m , maka akan dilakukan perhitungan nilai prediksi pada mata kuliah yang belum pernah diambil oleh *user* U_m . Berdasarkan pada hasil nilai prediksi ini, maka akan ditentukan mata kuliah terbaik yang akan direkomendasikan kepada



mahasiswa [1].

Gambar 1. Collaborative Filtering Process [1]

2.2.1 Item Based Collaborative Filtering

Item Based Collaborative Filtering merupakan suatu metode yang memberikan rekomendasi item dengan mencari item lain yang memiliki *Top-K similarity* item dengan item tersebut.

Kelebihan *Item Based Collaborative Filtering* [1]:

- *Scalability*

Dengan menggunakan metode ini model yang dihasilkan sangat kecil dibandingkan dengan dataset yang sebenarnya, meskipun diterapkan pada dataset yang lebih besar sekalipun, model yang dihasilkan masih sangat kecil untuk digunakan secara efisien.

- *Prediction Speed*

Lebih cepat dibandingkan dengan *User-Based* dikarenakan *dataset* yang digunakan dalam proses prediksi jauh lebih kecil.

- *Data Characteristics*

Lebih baik dipergunakan pada karakteristik data dimana terdapat perbedaan jumlah item yang besar antara *user* yang satu dengan *user* yang lainnya.

Kelemahan *Item Based Collaborative Filtering* [1]

- *Inflexibility*

Ketika terjadi penambahan item baru maka harus dilakukan perubahan pada *query*.

2.2.2 User Based Collaborative Filtering

User Based Collaborative Filtering merupakan suatu metode yang memberikan rekomendasi item dengan membandingkan semua *item* pada semua *user* aktif dengan *user* tertentu untuk mencari *Top-K similarity user* aktif dengan *user* tertentu.

Kelebihan *User Based Collaborative Filtering* [1]:

- Algoritma dapat digunakan pada banyak kondisi.
- Sangat mudah dalam melakukan update pada database dikarenakan pada algoritma ini prediksi selalu dilakukan berdasarkan pada semua *user* pada database.
- Lebih cocok dipergunakan pada karakteristik data dimana perbedaan jumlah item yang dimiliki oleh *user* yang satu dengan lainnya kecil atau bersifat merata.

Kelemahan *User Based Collaborative Filtering* [1]:

- Membutukan waktu yang lama.

Hal ini dikarenakan setiap terjadi penambahan data *user* baru, akan dilakukan update pada nilai prediksi.

User Based Collaborative Filtering dipilih berdasarkan karakteristik data dari mahasiswa beserta dengan mata kuliah yang diambil dimana terdapat jumlah sks mata kuliah yang hampir sama yang dimiliki oleh setiap mahasiswa dalam satu angkatan beserta dengan jumlah sks mata kuliah yang sama bagi setiap mahasiswa disetiap angkatan untuk mencapai kelulusan yaitu 144 sks.

2.3 Algoritma pada Collaborative Filtering

Dalam penggunaan dari metode pada *Collaborative Filtering*, terdapat algoritma yang dipergunakan sebagai basis dalam perhitungan pada metode *Collaborative Filtering* yang telah dipilih.

Dua macam algoritma yang dapat digunakan dalam menentukan nilai *similarity* diantaranya adalah:

2.3.1 Cosine Similarity

Cosine Similarity atau *Vector Based Similarity* merupakan algoritma dimana nilai *similarity* antara i dan j digambarkan sebagai suatu sudut yang terbentuk diantara 2 buah [5]. Pada algoritma ini apabila terdapat mahasiswa yang belum mengambil mata kuliah maka nilai pada mata kuliah tersebut akan dianggap sebagai nilai terendah yaitu 0.

$$\text{sim}(i, j) = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\|_2 * \|\vec{j}\|_2} \quad (1)$$

2.3.2 Adjusted Cosine Similarity

Adjusted Cosine Similarity merupakan pengembangan dari *Cosine Similarity* dimana terdapat pertimbangan akan adanya data / nilai yang kosong yang tidak dimiliki oleh *user* sehingga perlu dilakukan suatu normalisasi pada data untuk menjadikan data nilai kosong yang diasumsikan menggunakan 0 menjadi sebuah nilai tengah. Dalam studi kasus mata kuliah artinya terdapat mata

kuliah yang belum diambil oleh mahasiswa. Algoritma *Adjusted Cosine Similarity* ini dipilih dikarenakan dengan menggunakan algoritma ini maka nilai 0 yang mewakili nilai dari mata kuliah yang belum diambil oleh mahasiswa tidak lagi menjadi nilai terendah, namun menjadi nilai tengah dari perhitungan *similarity* sehingga dihasilkan hasil yang lebih akurat. Rumus dari *Adjusted Cosine Similarity* adalah sebagai berikut[4]:

$$sim(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_j)^2}} \quad (2)$$

Dimana:

- i merupakan mahasiswa yang nilainya akan diprediksi.
- j merupakan mahasiswa yang datanya digunakan untuk memprediksi.
- $sim(i, j)$ merupakan nilai *similarity* antara i dan j .
- $R_{u,i}$ merupakan nilai mata kuliah u dari mahasiswa i .
- $R_{u,j}$ merupakan nilai mata kuliah u dari mahasiswa j .
- \bar{R}_j merupakan nilai rata-rata dari semua mata kuliah dari mahasiswa j .
- \bar{R}_i merupakan nilai rata-rata dari semua mata kuliah dari mahasiswa i .

2.3.3 Nilai Prediksi

Setelah dilakukan perhitungan *similarity* untuk mencari *Top-K* mahasiswa yang memiliki *similarity* yang tertinggi dengan mahasiswa U_m maka langkah yang dilakukan adalah menghitung nilai prediksi pada mata kuliah yang belum diambil oleh mahasiswa U_m berdasarkan pada data nilai mata kuliah dari semua *Top-K* mahasiswa [2].

Rumus dari Nilai Prediksi dapat dinyatakan sebagai berikut[6]:

$$P_{u,i} = \bar{R}_a + \frac{\sum_{i \in U} (R_{u,i} - \bar{R}_i) * S_{i,j}}{\sum_{i \in U} |S_{i,j}|} \quad (3)$$

Dimana:

- i merupakan mahasiswa yang nilainya akan diprediksi.
- j merupakan mahasiswa yang datanya digunakan untuk memprediksi.
- $P_{u,i}$ adalah nilai prediksi mata kuliah yang dari mahasiswa i pada mata kuliah u .
- $i \in U$ merupakan *Top-K* mahasiswa yang memiliki *similarity* paling tinggi dengan mahasiswa.
- \bar{R}_a merupakan rata-rata dari semua mata kuliah mahasiswa yang dicari.
- $R_{u,i}$ merupakan nilai mata kuliah mahasiswa i terhadap mata kuliah u .
- \bar{R}_i merupakan nilai rata-rata dari semua mata kuliah dari mahasiswa u .
- $S_{i,j}$ adalah nilai *similarity*.

2.3.4 MAE (Mean Absolute Error)

MAE digunakan untuk menghitung rata-rata perbedaan secara mutlak antara nilai mata kuliah prediksi dengan nilai mata kuliah sebenarnya [3]. Jika $P_{u,i}$ adalah nilai prediksi mata kuliah dari mahasiswa u untuk mata kuliah i , maka $R_{u,i}$ adalah nilai mata kuliah yang sebenarnya.

Persamaan MAE dirumuskan sebagai berikut :

$$MAE = \frac{\sum_{\{u,i\}} |P_{u,i} - R_{u,i}|}{N} \quad (4)$$

Dimana :

- i merupakan mahasiswa yang nilainya akan diprediksi.
- $P_{u,i}$ adalah nilai prediksi mata kuliah yang dari mahasiswa i pada mata kuliah u .
- $R_{u,i}$ adalah nilai mata kuliah yang sebenarnya dari mahasiswa i .
- N adalah jumlah data.

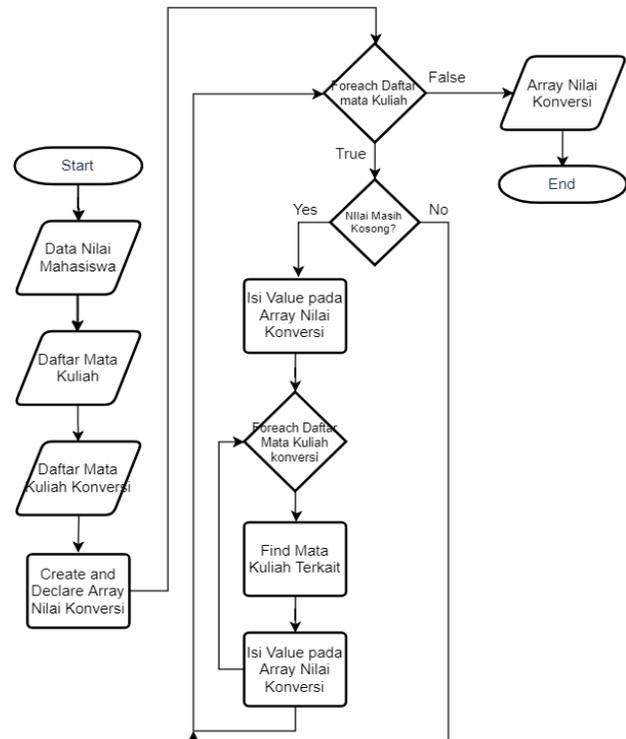
3. DESAIN SISTEM

3.1 Desain Flowchart

Desain dari *flowchart* dilakukan untuk menggambarkan proses yang terjadi pada suatu sistem yang dibuat. Dalam *flowchart* disajikan gambaran secara terperinci mengenai setiap *input*, proses dan *output* yang dihasilkan dari serangkaian proses yang terjadi dalam suatu sub sistem.

3.1.1 Flowchart Preprocessing

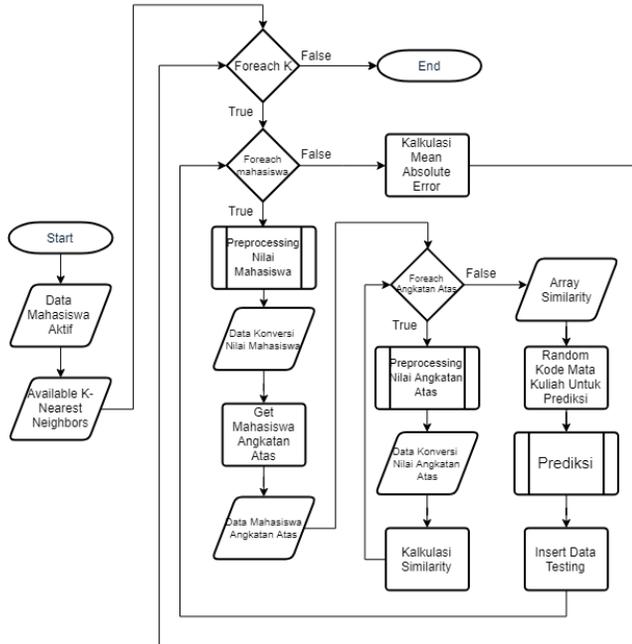
Flowchart ini menjelaskan proses *preprocessing* dimana dalam sub sistem ini dilakukan konversi pada nilai mahasiswa dari kurikulum yang sebelum 2017 menuju kepada kurikulum 2017. Dalam proses ini dihasilkan suatu array yang menampung nilai yang telah dikonversi menuju kurikulum 2017. Penggambaran alur dari *preprocessing* dapat dilihat pada Gambar 2.



Gambar 2. Flowchart Preprocessing

3.1.2 Flowchart Testing

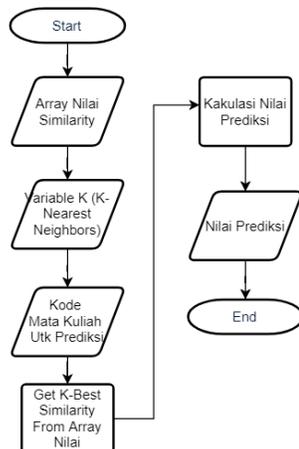
Dalam *flowchart* ini dijelaskan mengenai bagaimana pengujian dan penelitian pada sistem rekomendasi mata kuliah pilihan dapat dilakukan. Pengujian dilakukan terutama untuk menentukan *K-Nearest Neighbor* atau sejumlah *N* data dengan nilai *similarity* terbaik yang akan dipergunakan sebagai acuan dalam perhitungan nilai prediksi. Penggambaran alur dari *testing* dapat dilihat pada Gambar 3.



Gambar 3. Flowchart Testing

3.1.3 Flowchart Prediction

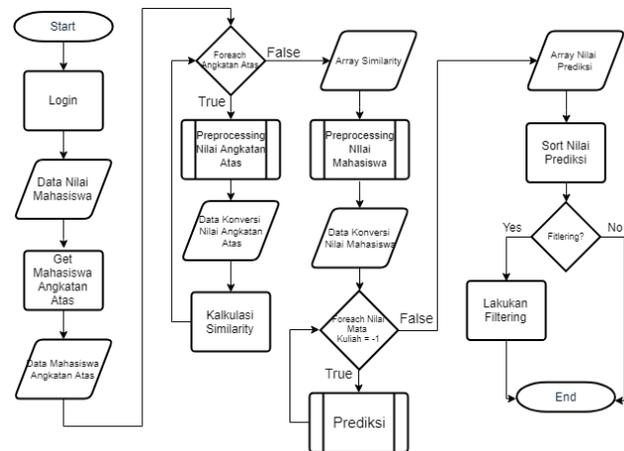
Flowchart prediction merupakan *flowchart* yang menggambarkan bagaimana suatu sistem dapat melakukan prediksi nilai dari suatu mata kuliah berdasarkan data yang ada. Penggambaran alur dari *prediction* dapat dilihat pada Gambar 4.



Gambar 4. Flowchart Prediction

3.1.4 Flowchart Recommendation

Flowchart recommendation menggambarkan alur *user* dari *login* hingga *user* mendapat rekomendasi mata kuliah pilihan pada aplikasi sistem rekomendasi. Penggambaran alur pada *recommendation* dapat dilihat pada Gambar 5.



Gambar 5. Flowchart Testing

4. IMPLEMENTASI SISTEM

Implementasi dilakukan pada komputer dengan spesifikasi:

- RAM : 8 GB, DDR3
- HDD : 1 TB
- CPU : Intel Core i5
- OS : Windows

Implementasi pengkodean sistem, menggunakan Bahasa pemrograman php. *Framework* yang digunakan untuk sistem ini adalah *Laravel*.

Database yang digunakan pada aplikasi ini adalah MySQL. Aplikasi ini dibuat dengan menggunakan arsitektur berbasis *Service Oriented*. Setiap fungsi pemograman dan *query database* akan diaplikasikan pada *Controller* dalam framework *laravel*. Pemanggilan fungsi pada *Controller* akan dilakukan dengan bantuan API (*Application Programming Interface*).

Pada setiap halaman terdapat *javascript* yang akan membantu sistem dalam melakukan *request* kepada *API*.

Setiap data memiliki akses dengan berbagai *HTTP Method*, antara lain:

- GET (Menerima data dalam *object*)
- POST (Memasukan data, INSERT)
- PUT (Mengubah data, UPDATE)
- DELETE (Menghapus data, DELETE)

5. ANALISA DAN PENGUJIAN

5.1 Pengujian pada Collaborative Filtering

Pengujian pada *Collaborative Filtering* akan dilakukan perbandingan performa dari *Cosine* dan *Adjusted Cosine* dalam sistem rekomendasi serta variabel *K-Nearest Neighbors* yang memberikan akurasi terbaik dalam memberikan rekomendasi.

5.1.1 Pengujian Cosine dan Adjusted Cosine Similarity

Pengujian ini dilakukan untuk membandingkan kesesuaian dari kedua algoritma dalam memberikan rekomendasi mata kuliah kepada mahasiswa. Pengujian akan dilakukan dengan menggunakan *Sampling* dan akurasi dari prediksi dalam

memberikan rekomendasi mata kuliah pilihan pada semua mahasiswa aktif.

5.1.1.1 Sampling

Pengujian menggunakan sampling berkaitan dengan kemungkinan pola sederhana yang ditemukan pada data nilai mahasiswa. Kemungkinan sampling yang ditemukan dan digunakan sebagai pengujian adalah:

- **Sample 1 (Data Nilai Sama)**

Pola data pada Sample 1 menggunakan pola dimana data nilai dari mahasiswa prediksi memiliki pola nilai yang sama dengan mahasiswa acuan. Pola dari *sample 1* dapat dilihat pada Tabel 1.

Tabel 1. Tabel Sample 1

NRP	NILAI MATA KULIAH				
	A	B	C	D	E
1(Prediksi)	B	C	B	C	?
2(Acuan)	B	C	B	C	B+

- **Sample 2 (Pola Data Sama Tetapi Nilai Berbeda)**

Pola data pada Sample 2 menggunakan pola dimana data nilai dari mahasiswa prediksi memiliki pola nilai yang sama dengan mahasiswa acuan akan tetapi terdapat perbedaan pada nilai yang dimiliki. Pola dari *sample 2* dapat dilihat pada Tabel 2.

Table 2. Tabel Sample 2

NRP	NILAI MATA KULIAH				
	A	B	C	D	E
1(Prediksi)	B	C	B	C	?
2(Acuan)	A	B	A	B	B

- **Sample 3 (Pola Bertolak Belakang)**

Pola data pada Sample 3 menggunakan pola dimana data nilai dari mahasiswa prediksi memiliki pola nilai yang bertolak belakang dengan mahasiswa acuan. Pola dari *sample 3* dapat dilihat pada Tabel 3.

Tabel 3. Tabel Sample 3

NRP	NILAI MATA KULIAH				
	A	B	C	D	E
1(Prediksi)	B	C	B	C	?
2(Acuan)	B	A	B	A	A

5.1.1.2 Pengujian Pada Data Nilai Mahasiswa Aktif

Selain menggunakan sampling, akan dilakukan prediksi pada seluruh mahasiswa aktif Informatika dan SIB untuk membandingkan akurasi dari Cosine dan Adjusted Cosine dalam memberikan rekomendasi.

5.1.2 Pengujian Pada K-Nearest Neighbors

Pengujian pada *K-Nearest Neighbors* dilakukan untuk mengetahui Variabel K yang memiliki akurasi terbaik sebagai acuan dalam melakukan prediksi nilai. Pengujian Variabel K akan dilakukan dengan *range* 5-20 pada 474 mahasiswa yang berstatus aktif pada

jurusan Informatika dan SIB. Hasil dari prediksi akan disimpan dalam Tabel Testing pada *database*.

5.2 Hasil Pengujian

Pengujian sistem rekomendasi dilakukan dengan 4 data, dari setiap data proses akan berjalan 5 kali dimana setiap iterasi aplikasi akan melakukan sampling dari 10% sampai 50% dengan rentan 10%.

5.2.1 Hasil Pengujian Cosine dan Adjusted Cosine

Pada bagian ini akan dijelaskan mengenai hasil dari pengujian perbandingan performa dari *Cosine Similarity* dan *Adjusted Cosine Similarity*.

5.2.1.1 Hasil Pengujian Sampling

➤ *Algorithm: Cosine Similarity*

Similarity: 1

Sample: 1 (Data Nilai Sama)

Result:

Tabel 4. Tabel Hasil Prediksi Pada Sample 1 dengan menggunakan Cosine

NRP	NILAI MATA KULIAH				
	A	B	C	D	E
1(Prediksi)	B	C	B	C	B+
2(Acuan)	B	C	B	C	B+

➤ *Algorithm: Adjusted Cosine Similarity*

Similarity: 1

Sample: 1 (Data Nilai Sama)

Result:

Tabel 5. Tabel Hasil Prediksi Pada Sample 1 dengan menggunakan Adjusted Cosine

NRP	NILAI MATA KULIAH				
	A	B	C	D	E
1(Prediksi)	B	C	B	C	B+
2(Acuan)	B	C	B	C	B+

➤ *Algorithm: Cosine Similarity*

Similarity: 0.998

Sample: 2 (Pola Data Sama Tetapi Nilai Berbeda)

Result:

Tabel 6. Tabel Hasil Prediksi Pada Sample 2 dengan menggunakan Cosine

NRP	NILAI MATA KULIAH				
	A	B	C	D	E
1(Prediksi)	B	C	B	C	A
2(Acuan)	A	B	A	B	A

➤ *Algorithm: Adjusted Cosine Similarity*

Similarity: 1

Sample: 2 (Pola Data Sama Tetapi Nilai Berbeda)

Result:

Tabel 7. Tabel Hasil Prediksi Pada Sample 2 dengan menggunakan Adjusted Cosine

NRP	NILAI MATA KULIAH				
	A	B	C	D	E
1(Prediksi)	B	C	B	C	B
2(Acuan)	A	B	A	B	A

- *Algorithm: Cosine Similarity*
Similarity: 0.943
Sample: 3 (Pola Bertolak Belakang)
Result:

Tabel 8. Tabel Hasil Prediksi Pada Sample 3 dengan menggunakan Cosine

NRP	NILAI MATA KULIAH				
	A	B	C	D	E
1(Prediksi)	B	C	B	C	A
2(Acuan)	B	A	B	A	A

- *Algorithm: Adjusted Cosine Similarity*
Similarity: -1
Sample: 3 (Pola Bertolak Belakang)
Result:

Tabel 9. Tabel Hasil Prediksi Pada Sample 3 dengan menggunakan Adjusted Cosine

NRP	NILAI MATA KULIAH				
	A	B	C	D	E
1(Prediksi)	B	C	B	C	B
2(Acuan)	B	A	B	A	A

Hasil Pengujian pada sample 1 menunjukkan dengan pola data yang sama, baik dari *Cosine Similarity* dan *Adjusted Cosine Similarity* memiliki nilai yang sama yaitu 1 yang merupakan nilai maksimal *similarity* dan menyatakan bahwa pola mahasiswa prediksi dan acuan memiliki pola yang sama sehingga apabila data nilai mahasiswa acuan dijadikan sebagai nilai prediksi maka nilai dari mahasiswa prediksi akan menunjukkan nilai yang sama dengan mahasiswa acuan yaitu "B+".

Hasil pengujian pada sample 2 menunjukkan dengan pola data sama namun nilai berbeda menunjukkan *similarity* yang sangat tinggi dari *Cosine* dengan nilai 0.943 dan nilai *similarity* sempurna dari *Adjusted Cosine* yaitu 1. Prediksi pada mata kuliah menggunakan *Cosine* menunjukkan nilai "A" sedangkan pada *Adjusted Cosine* menunjukkan nilai "B". Dalam hal ini *Cosine* meskipun memiliki nilai *similarity* yang tinggi, *Cosine* cenderung kurang baik dalam melakukan prediksi pada nilai. Hal ini dikarenakan pada *range* nilai dari mahasiswa prediksi adalah "B" sampai dengan "C" sehingga prediksi dari *Cosine* yaitu "A" merupakan prediksi yang berada di luar *range* dari nilai mahasiswa prediksi. Hasil pada *Adjusted Cosine* dengan *similarity* = 1 dan prediksi = "B" memiliki kualitas prediksi yang lebih baik dimana dalam hal ini "B" berada pada *range* nilai dari mahasiswa.

Hasil pengujian pada sample 3 dengan pola data yang saling bertolak belakang tetap menunjukkan *similarity* yang sangat tinggi pada *Cosine* yaitu 0.943 dan nilai *similarity* yang terendah pada *Adjusted Cosine* yaitu -1. Dalam hal ini akibat dari nilai *similarity* yang sangat tinggi dari *Cosine* membuat nilai prediksi cenderung berada pada kisaran yang sama dengan nilai pada mahasiswa acuan yaitu "A" sehingga pada akhirnya prediksi yang diberikan akan memiliki kecenderungan memberikan nilai yang sama yaitu "A". Sedangkan pada *Adjusted Cosine* dimana nilai *similarity* merupakan nilai *similarity* terendah yaitu -1 hasil dari prediksi tetap dalam *range* minimal dan maksimal dari nilai mahasiswa yaitu "B".

5.2.1.2 Hasil Pengujian Pada Data Nilai Mahasiswa Aktif

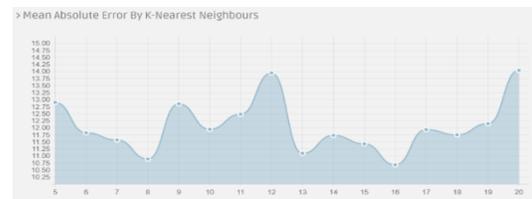
Tabel 10. Tabel Perbandingan Pengujian Pada Data Nilai Mahasiswa Aktif

	Cosine	Adjusted Cosine
Jumlah Data Testing	404	404
Threshold	0.1	0.1
Minimal Similarity	0.10014029468768496	0.10000002
Maximum Similarity	1	1
Mean Absolute Error	0.3075864930422574	0.1644052400076497
Accuracy	69.2413507 %	83.559476 %

Pada hasil pengujian seperti yang ditunjukkan oleh 0, *Adjusted Cosine* memiliki akurasi yang lebih baik dalam melakukan prediksi pada data nilai mahasiswa dengan **83.559476 %** dibandingkan **69.2413507 %** akurasi yang dimiliki oleh *Cosine*. Hal ini membantu dalam mendukung kesimpulan akan kesesuaian yang lebih baik dari *Adjusted Cosine Similarity* untuk dipergunakan sebagai algoritma sistem rekomendasi mata kuliah pilihan.

5.2.2 Hasil Pengujian K-Nearest Neighbors

Pada hasil pengujian pada variabel K dalam K-Nearest Neighbors akurasi terbaik ditunjukkan oleh Variabel K = 16 dimana dengan nilai variabel ini akurasi berada pada nilai yang maksimal yaitu **89.31422 %**. Visualisasi dari *Mean Absolute Error* dapat dilihat pada Gambar 6.



Gambar 6. Grafik Mean Absolute Error

6. KESIMPULAN

Setelah dilakukan perancangan sistem, pengimplementasian, dan pengujian terhadap aplikasi yang telah dibuat, dapat ditarik kesimpulan sebagai berikut:

- Algoritma Adjusted Cosine Similarity sangat sesuai untuk dipergunakan pada karakteristik data dengan nilai yang bersifat subjektif seperti nilai mata kuliah dan nilai *rating*.
- Algoritma Cosine Similarity sangat baik dipergunakan pada data yang tidak terdapat nilai yang bersifat subjektif seperti *similarity* antara text berdasarkan kemiripan kata dalam *text*.
- Algoritma Adjusted Cosine Similarity memiliki kelebihan yaitu data dari mahasiswa dengan kecenderungan nilai rata-rata tinggi juga dapat digunakan untuk memprediksi mahasiswa dengan kecenderungan memiliki nilai rata-rata yang rendah.
- Algoritma *Adjusted Cosine Similarity* dengan menggunakan *filter K-Nearest Neighbors* memiliki kelemahan yaitu prediksi dari mata kuliah hanya bisa dilakukan apabila dari *Top-K Similarity* yang ditentukan oleh K-Nearest Neighbors sudah pernah mengambil mata kuliah pilihan yang bersangkutan. Apabila dari *Top-K Similarity* yang ditentukan tidak ada satupun yang mengambil mata kuliah pilihan mata prediksi pada mata kuliah pilihan tidak dapat dilakukan.
- Konversi nilai mahasiswa dari kurikulum lama ke kurikulum baru masih terbatas pada tabel yang dibuat oleh *Admin* sehingga penambahan data acuan dalam melakukan konversi akan sangat bergantung pada pengetahuan *Admin* dalam menentukan daftar mata kuliah yang mempelajari *studi* yang sama di kurikulum yang berbeda.

Saran yang dapat diberikan untuk pengembangan penelitian ini lebih lanjut:

- Mempergunakan gabungan metode *Collaborative Filtering* dengan *Content Based Filtering* dalam memberikan

rekomendasi mata kuliah pilihan sehingga hasil dari rekomendasi tidak hanya bergantung pada satu faktor seperti pola nilai mahasiswa dalam *Collaborative Filtering* tetapi juga dapat memperhitungkan berbagai faktor lainnya seperti: jam mata kuliah yang tidak nyaman, minat, dosen pengajar dan beberapa faktor lainnya sebagai faktor yang menentukan nilai mata kuliah yang diambil oleh mahasiswa.

7. DAFTAR PUSTAKA

- [1] Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. 2001. Item Based Collaborative Filtering Recommendation Algorithms. *Proceedings of the 10th international conference on World Wide Web* (hal. 285-295). Hong Kong: ACM.
- [2] Yanti, N., Rahmi, R., & Ruliah. (2013). Penerapan Algoritma Collaborative Filtering Untuk Rekomendasi Games Hardware. *JUTISI*, 267-334.
- [3] Badriyah, T., Restuningtyas, I., & Setyorini, F. 2017. Sistem Rekomendasi Collaborative Filtering Berbasis *User* Algoritma Adjusted Cosine Similarity. *Prosiding Seminar Nasional ReTII ke-12 2017* (hal. 38-41). Yogyakarta: <https://journal.stnas.ac.id/>.
- [4] Putra, A., Mahmudy, W., & Setiawan, B. 2015. Sistem Rekomendasi Mata Kuliah Pilihan Mahasiswa dengan Content-based Filtering dan Collaborative Filtering. *Repository Jurnal Mahasiswa PTIIK Universitas Brawijaya*, vol 5, no 17.
- [5] Yao, G., & Cai, L. 2015. *User-based and Item-based Collaborative Filtering Recommendation Algorithms Design*. UCSD CSE.
- [6] Kaushik, S., & Tomar, P. 2015. Evaluation of Similarity Functions by using User based Collaborative Filtering approach in Recommendation Systems. *IJETT*, 194-200