

**UJI PERBANDINGAN *COLLABORATIVE FILTERING ALGORITHM*
(CFA) DENGAN *DISSYMETRICAL PERCENTAGE COLLABORATIVE
FILTERING ALGORITHM* (DSPCFA) PADA SISTEM
REKOMENDASI FILM BERDASARKAN PREDIKSI *RATING***



SKRIPSI

**Disusun Sebagai Salah Satu Syarat
untuk Memperoleh Gelar Sarjana Komputer
pada Departemen Ilmu Komputer/Informatika**

Disusun Oleh :

JOHAN EKO PURNOMO

24010314130107

DEPARTEMEN ILMU KOMPUTER/INFORMATIKA

FAKULTAS SAINS DAN MATEMATIKA

UNIVERSITAS DIPONEGORO

2019

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini :

Nama : Johan Eko Purnomo

NIM : 24010314130107

Judul : Uji Perbandingan *Collaborative Filtering Algorithm* (CFA) dengan *Dissymmetrical Percentage Collaborative Filtering Algorithm* (DSPCFA) pada Sistem Rekomendasi Film Berdasarkan Prediksi *Rating*

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang sepengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Semarang, 26 Juni 2019



Johan Eko Purnomo

24010314130107

HALAMAN PENGESAHAN

Judul : Uji Perbandingan *Collaborative Filtering Algorithm* (CFA) dengan
Dissymmetrical Percentage Collaborative Filtering Algorithm (DSPCFA)
pada Sistem Rekomendasi Film Berdasarkan Prediksi *Rating*

Nama : Johan Eko Purnomo

NIM : 24010314130107

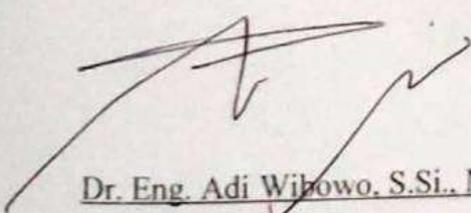
Telah diujikan pada sidang skripsi dan dinyatakan lulus pada tanggal **10 Mei 2019**.

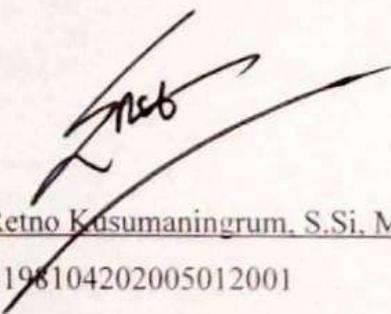
Semarang, 26 Juni 2019

Mengetahui,

a.n. Ketua Departemen Ilmu
Komputer/Informatika
Sekretaris

Panitia Penguji Skripsi
Ketua,


Dr. Eng. Adi Wibowo, S.Si., M.Kom
NIP. 198203092006041002


Dr. Retno Kusumaningrum, S.Si., M.Kom
NIP. 198104202005012001

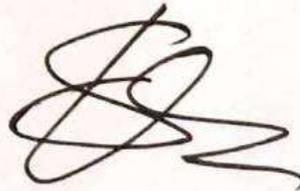
HALAMAN PENGESAHAN

Judul : Uji Perbandingan *Collaborative Filtering Algorithm* (CFA) dengan
Dissymmetrical Percentage Collaborative Filtering Algorithm (DSPCFA)
pada Sistem Rekomendasi Film Berdasarkan Prediksi *Rating*
Nama : Johan Eko Purnomo
NIM : 24010314130107

Telah diujikan pada sidang skripsi dan dinyatakan lulus pada tanggal **10 Mei 2019**.

Semarang, 26 Juni 2019

Menyetujui,
Dosen Pembimbing,



Sukmawati Nur Endah, S.Si, M.Kom.

NIP. 197805022005012002

ABSTRAK

Dalam era penyebaran informasi yang cepat khususnya dalam dunia industri hiburan perfilman, mencari film favorit baru untuk dinikmati berdasarkan ketertarikan pengguna menjadi sebuah tantangan dikarenakan banyaknya film yang tersedia. Sistem rekomendasi adalah salah satu solusi untuk memecahkan masalah tersebut, terlebih lagi apabila pengguna masih belum mengetahui kriteria atau judul film yang hendak dicari.

Algoritma sistem rekomendasi yang digunakan adalah *user-based collaborative filtering algorithm* (CFA) untuk memprediksi *rating* film yang akan direkomendasikan dengan langsung mengukur tingkat kemiripan/*similarity* antara pengguna target dengan pengguna lain tanpa memperhatikan banyaknya film yang sudah dirating oleh keduanya atau biasa disebut '*common items*'. Salah satu pendekatan berbeda dari algoritma CFA adalah algoritma *dissymmetrical percentage collaborative filtering algorithm* (DSPCFA) yang melibatkan banyaknya *common items* sebagai pertimbangan dari pengukuran similaritas. Algoritma DSPCFA menggunakan *threshold* yang berasal dari persentase terhadap jumlah *rating* yang dimiliki user target. *Threshold* tersebut akan menjadi batas minimal banyaknya *common items* yang harus terpenuhi pengguna lain sebelum diukur *similarity*-nya dengan pengguna target. Penelitian ini juga menggunakan dua metode pengukuran similaritas yaitu metode *pearson correlation similarity* dan metode *cosine similarity* sebagai perbandingan untuk mengetahui karakteristik dari masing-masing metode pengukuran.

Hasil pengujian menunjukkan bahwa algoritma DSPCFA menghasilkan nilai *error* yang lebih rendah dibandingkan algoritma CFA dengan penurunan sekitar 5% untuk metode evaluasi RMSE (*Root-mean Squared Error*) dan penurunan sekitar 7% menggunakan metode evaluasi MAE (*Mean Absolute Error*). Sedangkan untuk metode pengukuran yang diujikan menunjukkan bahwa metode *pearson correlation similarity* menunjukkan nilai *error* yang lebih rendah dibandingkan metode *cosine similarity*.

Kata Kunci : Sistem Rekomendasi, *Collaborative Filtering Algorithm*, *Dissymmetrical Percentage Collaborative Filtering Algorithm*, *Common Items*.

ABSTRACT

In era of rapid information transmissions, especially in the cinema industry, searching for a new favorite movie based on user interest is a challenge because of many movies that are available. The recommendation system is one kind of solution to solves this problem, especially if the user still does not know the criteria or the title of the movie he wants to find. User-based recommendation system is the

User-based collaborative filtering algorithm (CFA) is used to predict movie ratings which will be recommended by directly measuring the level of similarity between target users and other users regardless of the number of films that have been rated by both or commonly called 'common items'. One of the different approaches of the CFA algorithm is a dissymmetric percentage collaborative filtering algorithm (DSPCFA) algorithm which involves common items as a consideration of measuring similarity. The DSPCFA algorithm uses a threshold derived from the percentage of the number of ratings the target user has. The threshold will be the minimum limit for the number of common items that other users must meet before the similarity is measured with the target user. This study also uses two methods of measuring similarity, namely pearson correlation similarity and cosine similarity as a comparison to determine the characteristics of each measurement method.

The test results show that the DSPCFA algorithm produces a lower error value than the CFA algorithm with at least 5% error decrease using RMSE (*Root-mean Squared Error*) evaluation method and at least 7% error decrease using MAE (*Mean Absolute Error*) evaluation method. While test results for measurement method show that pearson correlation similarity produces lower error value than cosine similarity.

Keywords : Recommendation System, Collaborative Filtering Algorithm, Dissymmetrical Percentage Collaborative Filtering Algorithm, Common Items.

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa yang telah melimpahkan segala rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “Uji Perbandingan *Collaborative Filtering Algorithm* (CFA) dengan *Dissymmetrical Percentage Collaborative Filtering Algorithm* (DSPCFA) pada Sistem Rekomendasi Film Berdasarkan Prediksi *Rating*”.

Skripsi ini dibuat dengan tujuan sebagai salah satu syarat untuk memperoleh gelar sarjana komputer pada Departemen Ilmu Komputer/Informatika Fakultas Sains dan Matematika Universitas Diponegoro, Semarang.

Dalam pelaksanaan skripsi serta penyusunan dokumen skripsi ini, penulis menyadari banyak pihak yang membantu sehingga akhirnya dokumen ini dapat diselesaikan. Oleh karena itu, melalui kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Ibu Dr. Retno Kusumaningrum, S.Si, M.Kom selaku Kepala Departemen Ilmu Komputer/Informatika Fakultas Sains dan Matematika Universitas Diponegoro, Semarang.
2. Bapak Panji Wisnu Wirawan, ST, MT, selaku Koordinator Skripsi Departemen Ilmu Komputer/Informatika Fakultas Sains dan Matematika Universitas Diponegoro, Semarang.
3. Ibu Sukmawati Nur Endah, S.Si, M.Kom. selaku dosen pembimbing skripsi yang telah membantu dalam membimbing dan mengarahkan penulis hingga selesainya skripsi ini.
4. Semua pihak yang telah membantu kelancaran dalam penyusunan skripsi, yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa dokumen skripsi ini masih jauh dari sempurna. Oleh karena itu, saran dan kritik yang membangun sangat penulis harapkan. Akhir kata, semoga skripsi ini dapat bermanfaat bagi semua pihak.

Semarang, Mei 2019

Penulis

DAFTAR ISI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PENGESAHAN	iv
ABSTRAK.....	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL	xi
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah	4
1.3. Tujuan dan Manfaat.....	4
1.4. Ruang Lingkup	5
1.5. Sistematika Penulisan	5
BAB II LANDASAN TEORI	7
2.1. Penelitian Sebelumnya	7
2.2. Sistem Rekomendasi	8
2.3. <i>Collaborative Filtering Algorithm</i>	9
2.4. <i>Dissymmetrical-Percentage Collaborative Filtering Algorithm</i>	15
2.5. <i>Root Mean Squared Error (RMSE)</i> dan <i>Mean Absolute Error (MAE)</i>	17
2.6. <i>Dataframe Python</i>	19
2.7. <i>MovieLens</i>	19
BAB III METODOLOGI PENELITIAN	20
3.1. Garis Besar Penyelesaian Masalah	20
3.2. Pengumpulan Data.....	20
3.3. <i>Data Pre-processing</i>	24
3.4. Pembuatan <i>Dataframe Rating</i>	26
3.5. Pengukuran Similaritas.....	28
3.6. Pembuatan <i>Dataframe Similarity</i>	34
3.7. Rekomendasi film menggunakan CFA/DSPCFA	36
3.8. Evaluasi dan Interpretasi	41

BAB IV HASIL DAN PEMBAHASAN	45
4.1. Hasil Pengembangan Sistem.....	45
4.2. Skenario Pengujian Sistem.....	47
4.3. Hasil Skenario Pengujian	47
BAB V KESIMPULAN DAN SARAN	59
5.1. Kesimpulan	59
5.2. Saran.....	59
DAFTAR PUSTAKA	60

DAFTAR GAMBAR

Gambar 2.1 Contoh dasar algoritma <i>user-based collaborative filtering</i>	11
Gambar 2.2 Gambaran contoh <i>dataframe</i> pemetaan <i>user/film</i>	12
Gambar 3.1 Diagram Garis Besar Penyelesaian Masalah	21
Gambar 3.2 Flowchart algoritma CFA untuk Skenario 1A/B.....	29
Gambar 3.3 Flowchart algoritma DSPCFA untuk Skenario 2A/B.....	30
Gambar 3.4 Gambaran proses looping untuk mengisi <i>dataframe similarity</i>	35
Gambar 3.5 Flowchart untuk proses pengisian nilai rating prediksi untuk menghasilkan dataframe rating prediksi	40
Gambar 4.1 Gambaran proses beserta perangkat lunak yang digunakan dalam penelitian ..	46
Gambar 4.2 Grafik perbandingan nilai RMSE untuk Skenario 1 dalam persebaran nilai K ..	49
Gambar 4.3 Grafik perbandingan nilai MAE untuk Skenario 1 dalam persebaran nilai K	50
Gambar 4.4 Grafik perbandingan nilai RMSE untuk Skenario 2A (DSPCFA-Pearson)	54
Gambar 4.5 Grafik perbandingan nilai RMSE untuk Skenario 2B (DSPCFA-Cosine).....	56
Gambar 4.6 Grafik perbandingan nilai MAE untuk Skenario 2A (DSPCFA-Pearson)	57
Gambar 4.7 Grafik perbandingan nilai MAE untuk Skenario 2B (DSPCFA-Cosine).....	58

DAFTAR TABEL

Tabel 2.1	Penelitian terkait sistem rekomendasi film	7
Tabel 3.1	Pembagian Skenario 1 dan Skenario 2.....	20
Tabel 3.2	Tabel Atribut Dataset	22
Tabel 3.3	Gambaran <i>record</i> pada <i>file ratings.csv</i>	22
Tabel 3.4	Gambaran <i>record</i> pada <i>file movies.csv</i>	23
Tabel 3.5	Gambaran <i>record</i> pada <i>file links.csv</i>	23
Tabel 3.6	Gambaran struktur <i>movies.csv</i> setelah seleksi kolom	25
Tabel 3.7	Gambaran struktur <i>ratings.csv</i> setelah seleksi kolom.....	25
Tabel 3.8	Gambaran struktur <i>links.csv</i> setelah seleksi kolom	26
Tabel 3.9	Gambaran struktur data baru setelah proses <i>merge</i> data.....	26
Tabel 3.10	Gambaran <i>Dataframe rating</i> untuk 7 baris teratas.....	28
Tabel 3.11	Gambaran <i>Dataframe similarity</i> kosong dengan indeks <i>userId</i> dan kolom ' <i>similarity</i> '	34
Tabel 3.12	Contoh <i>dataframe similarity</i> untuk <i>userId</i> 671 dari 5 baris teratas untuk Skenario	36
Tabel 3.13	Contoh tabel <i>similarity</i> untuk <i>userId</i> 1 dengan $K=5$ (5 <i>user</i> terdekat)	37
Tabel 3.14	Gambaran <i>dataframe</i> prediksi dengan indeks <i>movieId</i> dan kolom ' <i>rating</i> ' ...	39
Tabel 3.15	Gambaran <i>dataframe</i> prediksi <i>rating</i> setelah melalui proses pengisian dan perhitungan nilai <i>rating</i> prediksi	41
Tabel 3.16	Contoh perbandingan <i>rating</i> asli dan <i>rating</i> prediksi menggunakan <i>cosine</i> <i>similarity</i> pada <i>userId</i> 1	42
Tabel 4.1	Data sampel <i>user</i> pengujian	47
Tabel 4.2	Hasil nilai error rating prediksi untuk Skenario 1 (CFA)	48
Tabel 4.2	Hasil nilai error rating prediksi untuk Skenario 2 (DSPCFA)	51
Tabel 4.4	Hasil nilai <i>error</i> minimum dari tiap skenario	56
Tabel 4.5	Hasil nilai <i>error</i> minimum algoritma DSPCFA terhadap CFA	56

BAB I

PENDAHULUAN

1.1. Latar Belakang

Saat ini arus informasi menjadi sangat cepat dan mudah dengan banyaknya jumlah informasi yang ada dan penting untuk memilah informasi yang diperlukan dan dibutuhkan menjadi fokus tersendiri dalam teknik pengolahan informasi. Banyaknya informasi juga terjadi pada dunia film, menurut *British Film Institute (BFI) film box office* yang diproduksi terus meningkat setiap tahunnya mulai tahun 2009 hingga tahun 2015 (British Film Institute, 2016). Pada tahun 2009 terdapat 503 film yang diproduksi dan sebanyak 759 film diproduksi pada tahun 2015 (British Film Institute, 2016). Pada era industri hiburan saat ini, kompleksitas informasi yang tinggi serta banyaknya informasi yang ada menyebabkan masalah yang berujung pada banyaknya waktu yang terbuang. Hal ini menjadi sangat menyusahakan untuk memilih film yang sesuai dengan selera para penikmat film. Informasi mengenai preferensi pengguna juga dibutuhkan oleh pihak produksi media film dan distributor untuk mendukung teknik pengiklanan mereka dan meningkatkan persaingan bisnis hiburan film. Salah satu cara untuk menangani hal tersebut adalah menggunakan sistem rekomendasi film.

Sistem rekomendasi adalah suatu teknologi yang didesain untuk mempermudah pengguna dalam menemukan suatu data yang mungkin sesuai dengan profil pengguna secara cepat dan dapat mengurangi jumlah informasi yang terlalu banyak (Vozalis & Margaritis, 2010). Sistem rekomendasi juga dapat diartikan sebagai program yang mencoba untuk merekomendasikan *items* baik berupa barang atau jasa untuk pengguna tertentu (individual atau bisnis) dengan memprediksi minat pengguna untuk sebuah item berdasarkan informasi terkait tentang item tersebut, pengguna, dan interaksi antar item dan pengguna (Bobadilla, 2013). Sistem rekomendasi film menampilkan prediksi film yang sesuai untuk pengguna, hal ini penting karena dapat mengurangi waktu yang diperlukan pengguna untuk mencari rekomendasi film yang sesuai dengan preferensi pengguna. Oleh karena itu, sistem rekomendasi sangat berguna untuk pengguna dan penyedia layanan hiburan untuk meningkatkan keefektifan cara pemasaran. Sistem rekomendasi juga terbukti untuk meningkatkan kualitas dan proses pengambilan keputusan (Isinkaye, 2015).

Sistem rekomendasi secara umum biasanya dibagi menjadi 3 teknik *filtering* yang dapat digunakan yaitu *content-based filtering*, *collaborative filtering*, *hybrid filtering*

(Isinkaye, 2015). Pada *content-based filtering*, difokuskan pada bagaimana memanfaatkan informasi yang terdapat pada profil pengguna, dan mengolah rekomendasi item berdasarkan preferensi dan informasi pada profil pengguna. Sementara untuk *collaborative filtering* menggunakan teknik yang berbeda, dimana pengguna yang memiliki kesamaan pada preferensi dan informasi profil dikelompokkan dan menggunakan informasi kolektif tersebut untuk membuat daftar rekomendasi ke pengguna. Sementara teknik *hybrid filtering* adalah sebuah kombinasi algoritma antara *content-based filtering* dan *collaborative filtering* yang secara implementasi lebih kompleks.

Dari ketiga pembagian teknik *filtering* tersebut *collaborative filtering* merupakan teknik yang paling matang dan paling banyak berkembang serta merupakan teknik yang paling banyak diimplementasikan. Beberapa permasalahan terkait dengan teknik *content-based filtering* antara lain keterbatasan analisis konten, *overspecialization*, dan kelengkapan data, sehingga teknik *content-based filtering* sangat bergantung dengan metadata atau informasi dari suatu item yang mencakup informasi yang lengkap dan terstruktur dari profil pengguna sebelum dapat diberikan rekomendasi (Isinkaye, 2015). Sementara untuk teknik *hybrid filtering* membutuhkan persyaratan yang cukup kompleks berupa profil pengguna dan data *rating* sehingga diperlukan sumber data yang lengkap dan terstruktur antara data pengguna, data film, dan data *rating*. Teknik *collaborative filtering* memungkinkan untuk memberikan rekomendasi hanya menggunakan data *rating* saja, serta dengan banyaknya data *rating* yang tersedia, memungkinkan teknik *collaborative filtering* mampu diimplementasikan secara lebih efektif dibandingkan teknik lainnya.

Penelitian ini menggunakan teknik *collaborative filtering* atau *collaborative filtering algorithm* (CFA) yang memprediksi *rating* film untuk direkomendasikan kepada pengguna dengan memanfaatkan *rating* film dari pengguna lainnya yang memiliki kesamaan perilaku/selera. Algoritma CFA membantu pengguna membuat pilihan berdasarkan pendapat pengguna lainnya (Balraj, 2016). Algoritma CFA sendiri memiliki kelebihan diantaranya terdapat peningkatan performa rekomendasi dan dapat menghasilkan rekomendasi yang lebih memiliki personalifikasi sehingga item rekomendasi lebih dinamis dan beragam (Bansari, 2017). Penggunaan algoritma CFA yang digunakan pada sistem ini membuat film yang direkomendasikan didasarkan pada kesamaan selera pengguna dan pemilihan film rekomendasi lebih didasarkan pada kesan yang sama yang hendak dicari pada sebuah film direkomendasikan.

Implementasi dari algoritma CFA seringkali menghasilkan hasil *similarity* antar *user* tanpa memperhatikan pengaruh dari banyaknya item/film yang dirating antar *user* tersebut, sehingga menghasilkan *user-user* yang kurang sesuai atau relevan dengan *user* target. Banyaknya item yang sama yang dirating oleh *user* berbeda dinamakan *common items*. Dengan pertimbangan bahwa ada kemungkinan kemiripan lebih besar apabila dua pengguna memiliki banyak kesamaan satu dengan lainnya, namun pendekatan normal pada pengukuran similaritas kurang memperhatikan pertimbangan tersebut, seringkali dua pengguna dengan sedikit *common items* bisa memiliki nilai *similarity* lebih besar dibandingkan dengan pengguna yang secara subjektif terlihat lebih mirip atau *similar*. Solusi untuk hal tersebut adalah dengan melibatkan banyaknya *common items* sebagai pertimbangan serta menjadi bagian dari metode pengukuran similaritas. Salah satu cara menggunakan *common items* sebagai variabel pengukuran similaritas adalah dengan menggunakan *threshold*. Terdapat tiga jenis variasi algoritma CFA yang menggunakan *threshold* berdasarkan basis pertimbangannya yaitu : *threshold* berbasis persentase, *threshold* berbasis nilai log, dan *threshold* berbasis nilai probabilitas. Ketiga variasi algoritma tersebut memiliki tingkat kompleksitasnya masing-masing, pada penelitian ini menggunakan *threshold* berbasis persentase karena memiliki tingkat kompleksitas yang cukup sederhana dan mudah diterapkan serta memiliki keefektifan yang cukup signifikan dibanding variasi algoritma yang lain (Wang, Zhang, & Rong, 2017). Penelitian ini membandingkan algoritma CFA konvensional dengan variasi dari algoritma berbasis CFA yang menggunakan *threshold* berbasis persentase dengan pendekatan tidak simetris yaitu *dissymmetrical-percentage collaborative filtering algorithm* (DSPCFA). Algoritma DSPCFA menggunakan presentase dari jumlah *rating* dari *user* target sebagai *threshold* untuk banyaknya *common items* minimal yang harus dimiliki antara *user* target dan *user* lain.

Penelitian ini menggunakan dua metode pengukuran similaritas yaitu metode *pearson correlation similarity* dan metode *cosine similarity*. Dua metode tersebut merupakan metode pengukuran yang hampir sama secara proses namun berbeda dalam penanganannya terhadap vektor masukan. Masing-masing metode pengukuran similaritas akan diterapkan pada kedua algoritma *filtering* yang dibandingkan sebelumnya sehingga pada penelitian ini memungkinkan untuk melihat perbedaan karakteristik dari kedua pengukuran similaritas tersebut.

Pengujian dilakukan membandingkan nilai *error* dari hasil *rating* prediksi yang dihasilkan algoritma CFA dibandingkan algoritma DSPCFA sekaligus membandingkan karakteristik dari dua metode pengukuran similaritas. Pengujian tersebut berupaya mengetahui pengaruh penggunaan *common items* sebagai *threshold* pengukuran similaritas terhadap hasil prediksi *rating*, sehingga dengan hasil tersebut dapat memberikan wawasan yang bisa digunakan untuk perkembangan sistem rekomendasi kedepannya untuk meningkatkan kualitas *output* sistem rekomendasi.

1.2. Rumusan Masalah

Berdasarkan permasalahan yang telah disampaikan pada latar belakang, rumusan masalah adalah mengevaluasi *rating* prediksi yang dihasilkan oleh algoritma *collaborative filtering algorithm* (CFA) dan algoritma *dissymmetrical-percentage collaborative filtering algorithm* (DSPCFA) pada sistem rekomendasi film serta membandingkan karakteristik berdasarkan evaluasi nilai *error* dari dua metode pengukuran similaritas yaitu metode *pearson correlation similarity* dan metode *cosine similarity*.

1.3. Tujuan dan Manfaat

Tujuan dari penelitian ini adalah mengetahui perbandingan nilai *error* dari *rating* prediksi yang dihasilkan algoritma CFA dan algoritma DSPCFA pada sistem rekomendasi film menggunakan dua metode pengukuran similaritas yaitu metode *pearson correlation similarity* dan metode *cosine similarity*, sehingga terlihat pengaruh dalam penggunaan variabel *threshold* menggunakan *common items* dan perbandingan metode pengukuran similaritas pada *rating* prediksi.

Adapun manfaat yang diharapkan dari penelitian ini adalah wawasan yang didapat dari membandingkan karakteristik dan evaluasi *error* dari dua metode pengukuran similaritas yang telah disebutkan, memperhatikan pengaruh penggunaan *common items* serta memperhatikan kelebihan dan kelemahan dari masing-masing metode pengukuran similaritas untuk meningkatkan kualitas hasil rekomendasi.

1.4. Ruang Lingkup

Ruang Lingkup yang dibahas dalam penelitian ini adalah:

1. *Dataset* yang digunakan terdiri dari 100.004 *rating* film dari 9.125 film oleh 671 pengguna dengan waktu *rating* berkisar dari 09 Januari 1995 sampai dengan 16 Oktober 2016, yang didapatkan dari *movielens.org*
2. Sistem memberikan output berupa daftar *n-item* yang merupakan rekomendasi yang dihasilkan dengan *n default* adalah 10 item dengan minimal item untuk pengguna mendapatkan rekomendasi adalah 20 item yang sudah *dirating*
3. Sistem yang dibangun berbasis *web* dengan *framework flask* dan *python*

1.5. Sistematika Penulisan

Penulisan dokumen ini terdiri dari 5 (lima) bab untuk memberikan gambaran yang jelas dan terurut mengenai penyusunan Sistem, yaitu :

BAB I PENDAHULUAN

Bab ini membahas tentang latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup serta sistematika penulisan mengenai pengembangan sistem.

BAB II LANDASAN TEORI

Bab ini berisikan tentang teori-teori yang mendukung dalam merancang sistem. Tinjauan pustaka yang digunakan dalam penyusunan penelitian ini meliputi Sistem Rekomendasi, *Collaborative Filtering Technique*, *Rating Prediction*, *User-based Similarity Measure (Pearson Correlation Similarity dan Cosine Similarity)*, *K-Nearest Neighbor*, *Root Mean Squared Error*, dan *Mean Absolute Error*.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan tentang tahapan yang dilakukan pada penelitian ini. Tahapan tersebut meliputi *Preprocessing*, Menghitung nilai *similarity*, Memperhatikan *common items* dalam perhitungan nilai *similarity*, Seleksi menggunakan Metode *K-Nearest Neighbor*, Menghitung *rating* prediksi, Menentukan *N-film* rekomendasi teratas, Evaluasi Kinerja Sistem.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menguraikan hasil skenario eksperimen dan analisa pada penelitian dan penjelasan mengenai semua skenario eksperimen dan analisa dari setiap hasil eksperimen yang telah dilakukan.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan berdasarkan uraian pada bab-bab sebelumnya dan saran untuk pengembangan sistem selanjutnya.