

ANALISIS DAN IMPLEMENTASI METODE ITEM-BASED COLLABORATIVE FILTERING UNTUK SISTEM REKOMENDASI KONSENTRASI DI STMIK STIKOM BALI

I Wayan Jepriana¹, Shofwan Hanief²

^{1,2}Program Studi Sistem Informasi, Institut Teknologi dan Bisnis STIKOM Bali
Denpasar, Indonesia

e-mail: jepriana@stikom-bali.ac.id¹, hanief@stikom-bali.ac.id²

Abstrak

Pada penelitian ini dilakukan pengembangan sistem rekomendasi dengan metode *item-based collaborative filtering* untuk pemilihan konsentrasi di STMIK STIKOM Bali. Penelitian ini menggunakan data nilai mahasiswa Program Studi Sistem Komputer dan Sistem Informasi. Data mata kuliah digunakan sebagai item, data mahasiswa digunakan sebagai *user* dengan nilai untuk mata kuliah yang telah ditempuh digunakan sebagai *rating*. Evaluasi dilakukan terhadap efektivitas dan relevansi rekomendasi yang diberikan oleh sistem. Efektivitas diukur dengan *precision*, dan relevansi diukur menggunakan *recall*. Berdasarkan hasil pengujian, sebesar 98% dari rekomendasi yang diberikan oleh sistem terbukti menghasilkan indeks prestasi mata kuliah konsentrasi yang lebih besar atau sama dengan 2,75 pada kedua data program studi. Rekomendasi yang diberikan oleh sistem relevan terhadap 76% mahasiswa yang memperoleh indeks prestasi untuk mata kuliah konsentrasi lebih besar atau sama dengan 2,75 pada Program Studi Sistem Komputer dan 73% pada Program Studi Sistem Informasi. Dengan demikian disimpulkan sistem dapat memberikan rekomendasi yang cukup baik pada kedua program studi, namun masih perlu ditingkatkan dari sisi relevansi rekomendasi. Faktor yang berpengaruh terhadap rekomendasi yang dihasilkan sistem adalah nilai kemiripan yang digunakan untuk menghitung prediksi. Disarankan untuk melakukan pengembangan kombinasi metode yang dapat meningkatkan kualitas dan relevansi dari rekomendasi. Pengembangan dapat mempertimbangkan penggunaan algoritme genetika untuk memperbaiki hasil pengukuran nilai kemiripan yang diharapkan dapat meningkatkan kualitas dan relevansi rekomendasi yang dihasilkan.

Kata kunci: sistem rekomendasi, *collaborative filtering*, konsentrasi

Abstract

In this research, a recommendation system with an item-based collaborative filtering method was developed to select a concentration in STMIK STIKOM Bali. This study has used student data scores in the Computer Systems and Information Systems Study Program. Course data is used as an item, and student data is used as a user with grades for courses taken as a rating. An evaluation is carried out on the effectiveness and relevance of the recommendations given by the system. Effectiveness is measured by precision and relevance is measured using recall. Based on the results of the test, 98% of the recommendations given by the system proved to produce a concentration index of achievement of subjects greater than or equal to 2.75 in both study program data. The recommendations given by the system are relevant to 76% of students who obtain an achievement index for a concentration course greater than or equal to 2.75 in the Computer Systems Study Program and 73% in the Information Systems Study Program. Based on these results the system can provide fairly good recommendations on both study programs, but can still be improved in terms of the relevance of the recommendations. The factor that influences the recommendations produced by the system is the similarity value used to calculate predictions. It is recommended to develop a combination of methods that can improve the quality and relevance of the

recommendations. Future development can consider the use of genetic algorithms to improve the similarity value that are expected to improve the quality and relevance of the resulting recommendations.

Keywords : *recommendation system, collaborative filtering, concentration*

PENDAHULUAN

Pemilihan konsentrasi adalah suatu tahap yang wajib dilakukan oleh mahasiswa STIKOM Bali guna menentukan fokus spesialisasi bidang yang ingin dipelajari. Konsentrasi akan menentukan mata kuliah yang wajib di tempuh pada semester berikutnya. Program Studi S-1 Sistem Komputer di STIKOM Bali memiliki dua konsentrasi penjurusan sedangkan Program Studi S-1 Sistem Informasi memiliki empat konsentrasi penjurusan. Setiap konsentrasi memiliki tiga mata kuliah yang dijadikan prasyarat. Kewajiban agar mahasiswa memilih satu konsentrasi, sementara nilai dari mata kuliah prasyarat mencukupi untuk memilih lebih dari satu konsentrasi sering menjadi permasalahan bagi mahasiswa.

Pilihan jurusan yang tepat dapat membuat suatu perbedaan besar di perguruan tinggi. Kecocokan antara jurusan dengan ketertarikan dan kecenderungan kemampuan diperlukan oleh mahasiswa. Hal itu diperlukan karena mahasiswa memiliki kemampuan berpikir yang berbeda-beda dan minat yang berbeda-beda pula untuk berbuat sesuatu [1]. Mahasiswa bisa kesulitan atau bahkan menemui kegagalan pada suatu mata kuliah tertentu dan akan berdampak pada nilai karena konsentrasi yang tidak sesuai dengan minat. Pemilihan jurusan dan motivasi belajar mempunyai pengaruh secara parsial terhadap hasil belajar siswa [2]. Peminatan berpengaruh terhadap prestasi belajar [3].

Sistem rekomendasi adalah suatu teknik dan perangkat lunak yang dapat memberikan suatu anjuran atau rekomendasi yang sekiranya bermanfaat bagi pengguna di mana anjuran tersebut dapat membantu dalam proses pengambilan keputusan [4]. Sistem rekomendasi juga disebut sebagai sebuah sistem yang dapat membantu pengguna dalam mengatasi informasi yang berlebih dengan memberikan rekomendasi spesifik

bagi pengguna dan diharapkan rekomendasi tersebut bisa memenuhi keinginan dan kebutuhan pengguna [5]. Sistem rekomendasi saat ini sudah sangat berkembang dan digunakan dalam banyak aplikasi, baik itu aplikasi untuk perusahaan maupun untuk perorangan atau individu. Sistem rekomendasi dapat membantu pengguna dalam memilih kebutuhan berdasarkan preferensinya [6].

Collaborative Filtering (CF) adalah teknik sistem rekomendasi yang banyak digunakan dan mungkin merupakan yang paling umum. *Collaborative filtering* memanfaatkan informasi *rating* dari beberapa pengguna untuk memprediksi *rating* item untuk pengguna tertentu [7]. *Collaborative filtering* dibagi menjadi dua macam yaitu *user-based collaborative filtering* dan *item-based collaborative filtering* [8]. Algoritme *item-based* memberikan kinerja yang lebih baik daripada algoritme *user-based*, sementara pada saat yang sama algoritme *item-based* juga memberikan kualitas prediksi yang lebih baik dari pada algoritme *user-based* [9]. Metode pengukur kemiripan *adjusted cosine similarity* dapat menghasilkan kualitas prediksi terbaik berdasarkan galat prediksi yang dihasilkan ketika dibandingkan dengan metode pengukur kemiripan *cosine similarity* dan *adjusted cosine similarity*. Jumlah *neighborhood* yang berpengaruh signifikan untuk membuat prediksi adalah 20 pada kasus konsentrasi penjurusan [10]. Pada penelitian dari [11] metode *collaborative filtering* digunakan untuk pengembangan sistem rekomendasi laptop.

Berdasarkan permasalahan pemilihan konsentrasi yang ada di STIKOM Bali, pada penelitian ini dilakukan pengembangan sistem rekomendasi dengan metode *item-based collaborative filtering* untuk pemilihan konsentrasi di STMIK STIKOM Bali. Metode *item-based collaborative filtering* dipilih karena sesuai dengan konsep rekomendasi

konsentrasi yang di landasi kemiripan suatu mata kuliah prasyarat dengan mata kuliah konsentrasi. Mata kuliah dapat digunakan sebagai item dan mahasiswa yang mengambil suatu mata kuliah dapat digunakan sebagai *user* dalam metode *collaborative filtering*. Keluaran dari penelitian ini berupa analisa hasil sistem rekomendasi dengan metode *item-based collaborative filtering* yang diterapkan pada kasus pemilihan konsentrasi di STMIK STIKOM Bali. Tujuan dari penelitian ini adalah untuk mengetahui efektifitas dan relevansi dari rekomendasi yang dihasilkan metode *item-based collaborative filtering* untuk pemilihan konsentrasi di STMIK STIKOM Bali.

Collaborative Filtering

Collaborative filtering merupakan penerapan algoritmis dari upaya meniru bagaimana manusia saling bertukar rekomendasi dengan temannya. Di praktik nyata tidak semua rekomendasi teman cocok dengan selera, sehingga dapat dipilah antara teman mana yang memiliki selera serupa dan teman mana yang selernya tidak serupa untuk mendapatkan hasil rekomendasi yang paling sesuai [12]. Metode *collaborative filtering* didasarkan pada asumsi bahwa pengguna serupa lebih menyukai item serupa atau pengguna mengekspresikan preferensi serupa untuk item serupa. Alih-alih melakukan pengindeksan konten atau analisis konten, sistem *collaborative filtering* sepenuhnya bergantung pada rating dari anggota komunitas yang berpartisipasi. Metode *collaborative filtering* dikategorikan menjadi dua kelas umum, yaitu *model-based* dan *memory-based* [13].

Algoritme berbasis model (*model-based*) menggunakan data dasar untuk mempelajari model probabilistik, seperti model *cluster* atau model jaringan Bayesian, menggunakan statistik dan teknik *machine learning*. Sementara metode berbasis memori (*memory-based*), menyimpan informasi preferensi mentah di memori komputer dan mengaksesnya sesuai kebutuhan untuk menemukan pengguna atau item serupa dan membuat prediksi. Metode *collaborative filtering*

berbasis memori dapat dibagi lagi menjadi dua kelompok, yaitu metode berbasis pengguna (*user-based*) dan berbasis item (*item-based*). Di satu sisi, metode berbasis pengguna mencari pengguna (juga disebut "tetangga") yang serupa dengan pengguna aktif dan menghitung perkiraan *rating* sebagai rata-rata bobot dari *rating neighborhood* pada item yang diinginkan. Di sisi lain, metode berbasis item mencari item serupa untuk pengguna aktif.

Item-based Collaborative Filtering

Item-based collaborative filtering merupakan kebalikan dari metode *user-based*. Bila pada *user-based collaborative filtering* prediksi rekomendasi didasarkan oleh kemiripan antar pengguna, pada *item-based* prediksi untuk rekomendasi didasarkan oleh kemiripan antar item. Pendekatan berbasis item melihat ke dalam rangkaian item yang telah diberikan *rating* oleh pengguna target dan menghitung seberapa mirip item sasaran tersebut dengan item i dan kemudian memilih k item yang paling mirip $\{i_1, i_2, \dots, i_k\}$. Pada saat yang sama, nilai kemiripan $\{s_{i1}, s_{i2}, \dots, s_{ik}\}$ dari masing-masing item tersebut juga dihitung. Begitu item yang paling mirip ditemukan, prediksi dihitung dengan mengambil *weighted average* dari peringkat yang telah diberikan oleh pengguna target pada k item yang paling mirip. Terdapat dua aspek pada metode *item-based collaborative filtering*, yaitu perhitungan kemiripan dan komputasi prediksi [9].

Adjusted-cosine Similarity

Salah satu langkah penting dalam metode *item-based collaborative filtering* adalah menghitung kemiripan antara item dan kemudian memilih item yang paling mirip untuk membuar prediksi. Ide dasar dalam perhitungan kemiripan antara dua item i dan j adalah menemukan pengguna yang telah menilai kedua item (co-rated) ini dan kemudian menerapkan teknik penghitungan kemiripan untuk menentukan kesamaan s_{ij} . Komputasi nilai kemiripan menggunakan *cosine similarity* dalam kasus berbasis item memiliki satu kelemahan penting, perbedaan skala rating antara pengguna yang berbeda tidak

diperhitungkan. Metode *adjusted-cosine similarity* memperbaiki kekurangan ini dengan mengurangi nilai *rating* dengan rata-rata *rating* yang diberikan oleh masing-masing pengguna pada kasus *co-rated*. Untuk menghitung nilai kemiripan antara item i dan j dilakukan dengan menggunakan Persamaan (1), di mana \bar{R}_u adalah rata-rata *rating* pengguna ke- u .

$$\text{sim}(i, j) = \frac{\sum (R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum (R_{u,j} - \bar{R}_u)^2}} \quad (1)$$

Komputasi Prediksi

Langkah paling penting dalam sistem *collaborative filtering* adalah menghasilkan keluaran berupa prediksi *rating*. Begitu kumpulan item yang paling mirip terisolasi berdasarkan ukuran kemiripan, langkah selanjutnya adalah melihat *rating* yang telah diberikan oleh pengguna target dan menggunakan teknik untuk menghasilkan prediksi. Pada tahap ini digunakan teknik *Item-based nearest neighbor algorithms*. *Item-based nearest neighbor algorithms* adalah teknik untuk menghasilkan prediksi berdasarkan kemiripan antar item [8]. Prediksi *rating* untuk sebuah item harus didasarkan pada *rating* yang telah diberikan pengguna untuk item yang serupa. Persamaan (2) merupakan algoritme prediksi untuk pendekatan *item-based*. Prediksi *rating* yang diberikan oleh user u untuk item i terdiri dari *weighted sum rating user u* untuk item yang paling mirip dengan item i .

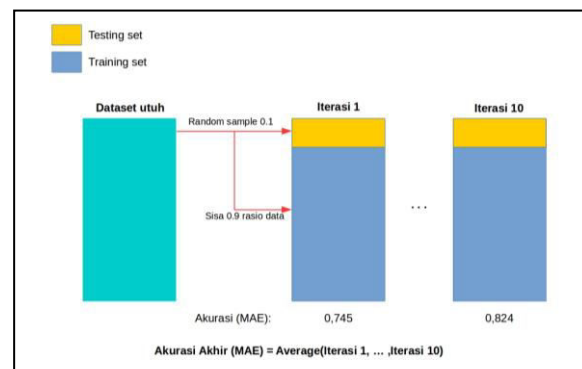
$$\text{pred}(u, i) = \frac{\sum \text{itemSim}(i, j) \cdot r_{uj}}{\sum \text{itemSim}(i, j)} \quad (2)$$

K-Fold Cross Validation

K-fold cross validation digunakan untuk menghitung akurasi prediksi suatu sistem. *K-fold cross validation* membagi data menjadi k buah segmen yang memiliki rasio yang sama atau hampir sama. Dilakukan *training* dan validasi sebanyak k kali dengan tiap perulangannya mengambil satu segmen berbeda sebagai data testing atau validasi dan $k-1$ segmen sisanya sebagai data *training* untuk kemudian

diambil nilai rata-rata dari hasil tiap iterasi [14].

Gambar 1 adalah ilustrasi dari *10-fold cross validation*. *K-fold cross validation* memanfaatkan metode *hold-out* data sebagai data testing, akan tetapi mengubahnya menjadi lebih efisien dengan melakukan k buah perulangan pada k buah *hold-out* data yang berbeda. Pada setiap perulangan berakhir, data tersebut dikembalikan ke data set awal. Ini memastikan setiap sudut data set mendapat kesempatan untuk ikut diuji.



Gambar 1. Ilustrasi *10-fold cross validation*

Mean Absolute Error

Untuk menilai performa suatu sistem rekomendasi dibutuhkan suatu metode dan perhitungan yang dapat mengukur tingkat kualitas prediksi yang dihasilkan oleh sistem. *Mean Absolute Error (MAE)* adalah salah satu metode statistika yang digunakan dalam mengevaluasi akurasi suatu sistem dengan membandingkan nilai hasil prediksi dengan nilai sesungguhnya pada data uji. Semakin rendah nilai MAE, semakin akurat prediksi yang telah dihasilkan. MAE menghitung penyimpangan nilai prediksi dari nilai sesungguhnya, untuk setiap pasang nilai prediksi dan nilai sesungguhnya (p_i dan q_i) dituliskan dalam Persamaan (3).

$$\text{MAE} = \frac{\sum_{i=1}^N |p_i - q_i|}{N} \quad (3)$$

Evaluasi

Evaluasi dilakukan untuk mengetahui efektivitas dan relevansi dari rekomendasi yang diberikan oleh sistem. Efektivitas rekomendasi dihitung dengan

menggunakan *precision*. Sementara relevansi rekomendasi dihitung menggunakan *recall*. Pada penelitian ini nilai *precision* dan *recall* didasarkan pada hasil *confusion matrix*, seperti pada Tabel 1.

Tabel 1. Confusion matrix

| Matriks | Aktual | | |
|-------------|---------|-------|----|
| | True | False | |
| Rekomendasi | Positif | TP | FP |
| | Negatif | TN | FN |

Keterangan:

- TP : Benar positif
- FP : Salah positif
- TN : Benar negatif
- FN : Salah negatif

Precision dan *recall* merupakan salah satu pengujian dasar dan paling sering digunakan dalam penentuan efektivitas *information retrieval system* maupun *recommendation system*. *True positive* (TP) pada *information retrieval* merupakan item relevan yang dihasilkan oleh sistem. Sementara *false negative* (FN) merupakan semua item relevan yang tidak dihasilkan oleh sistem. Istilah *positive* dan *negative* mengacu pada prediksi yang dilakukan oleh sistem. Sedangkan istilah *true* dan *false* mengacu pada prediksi yang dilakukan oleh pihak luar atau pihak yang melakukan observasi [15].

Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. *Precision* dihitung dengan Persamaan (4).

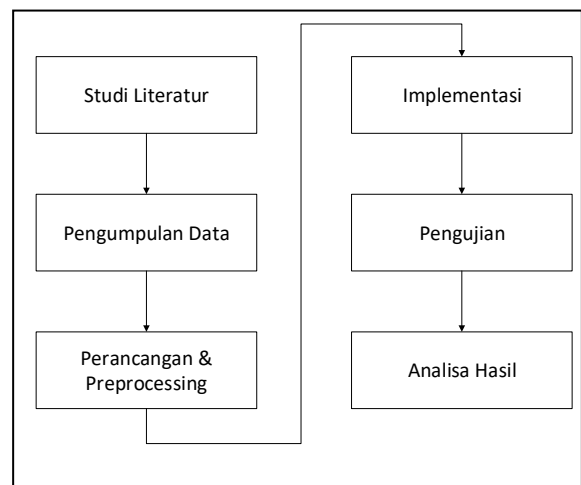
$$Precision = \frac{TP}{TP + FP} \quad (4)$$

Recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. *Recall* digunakan sebagai ukuran seberapa relevan rekomendasi yang dihasilkan oleh sistem. Dalam evaluasi *information retrieval system*, *recall* dihitung dengan Persamaan (5).

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

METODE

Pada penelitian ini yang menjadi obyek penelitian adalah STM IK STIKOM Bali atau yang sekarang telah berubah bentuk menjadi Institut Teknologi dan Bisnis STIKOM Bali yang berlokasi di Denpasar. Data yang digunakan diperoleh dari bagian Pusat Komputer dan Jaringan. Data yang digunakan pada penelitian ini adalah data nilai dari Program Studi Sistem Komputer dan Sistem Informasi. Langkah-langkah atau penelitian dilakukan sesuai dengan Gambar 2.



Gambar 2. Tahapan penelitian

Studi Literatur

Dilakukan studi literatur mengenai referensi yang sekiranya berhubungan dengan penelitian yang dilakukan. Pada tahap ini dipelajari literatur tentang konsep sistem rekomendasi dengan metode *collaborative filtering* khususnya *item-based collaborative filtering* untuk kasus konsentrasi dari berbagai buku, jurnal, dan artikel lainnya. Dari hasil studi literatur diketahui bahwa metode *item-based collaborative filtering* memiliki performa lebih baik dari pada metode *user-based collaborative filtering*. Metode pengukur kemiripan yang dapat menghasilkan galat prediksi paling rendah adalah metode pengukur kemiripan *adjusted-cosine similarity*. Dalam penelitian ini data nilai untuk suatu mata kuliah digunakan sebagai *rating* untuk suatu item, sementara mahasiswa digunakan sebagai *user*.

Pengumpulan Data

Berdasarkan studi literatur yang telah dilakukan, data yang dibutuhkan pada penelitian ini adalah data mahasiswa, data mata kuliah, dan data nilai dari mahasiswa yang telah mengambil konsentrasi. Adapun data yang digunakan adalah data mahasiswa Program Studi Sistem Komputer dan Program Studi Sistem Informasi angkatan 2015. Dipilih data mahasiswa angkatan 2015 karena sebagian besar mahasiswa angkatan 2015 telah memilih konsentrasi. Hal tersebut diperlukan untuk proses validasi rekomendasi yang diberikan oleh sistem.

Data yang digunakan pada penelitian ini adalah data primer yang diambil dari data internal nilai mahasiswa di STMIK STIKOM Bali angkatan 2015. Pada penelitian ini data direpresentasikan ke dalam bentuk matriks Mahasiswa-Mata kuliah dengan nilai *rating* diambil dari skala nilai kualitatif (1-4) yang di diperoleh mahasiswa untuk mata kuliah tertentu. Tabel 2 merupakan ilustrasi nilai mahasiswa yang direpresentasikan dalam bentuk matriks mahasiswa terhadap mata kuliah. Terdapat 5 orang mahasiswa dan 5 mata kuliah, di mana setiap mahasiswa telah memperoleh nilai setidaknya untuk 3 mata kuliah dari 5 mata kuliah yang ada.

Tabel 2. Ilustrasi representasi data nilai

| | MK_1 | MK_2 | MK_3 | MK_4 | MK_5 |
|-------|------|------|------|------|------|
| Mhs_1 | 3 | 3,5 | 3 | | |
| Mhs_2 | 4 | 3 | 3,5 | | |
| Mhs_3 | 3,5 | 4 | 3 | | |
| Mhs_4 | 4 | 3 | 3,5 | | |
| Mhs_5 | 2,5 | 2,75 | 3 | | |

Dari 382 orang mahasiswa program studi sistem komputer, terdapat 344 orang yang sudah mengambil konsentrasi. Pada program studi sistem komputer terdapat 65 pilihan mata kuliah yang terdiri dari 5 mata

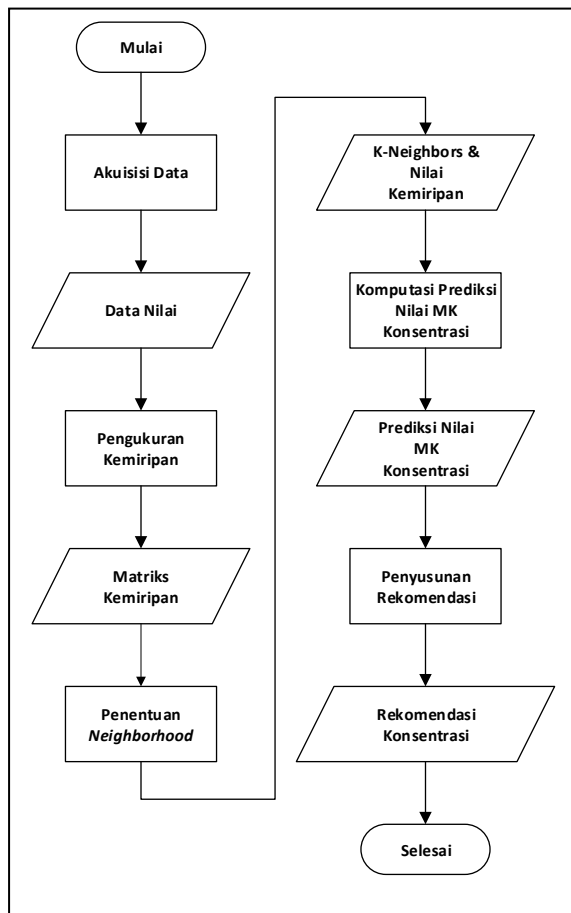
kuliah prasyarat dan 10 mata kuliah konsentrasi. Setiap konsentrasi memiliki 3 mata kuliah prasyarat dan 5 mata kuliah konsentrasi yang bersifat wajib.

Pada data program studi sistem informasi angkatan 2015, terdapat 687 orang mahasiswa dengan 673 orang sudah mengambil konsentrasi. Pada program studi sistem informasi terdapat 75 pilihan mata kuliah yang terdiri dari 11 mata kuliah prasyarat dan 27 mata kuliah konsentrasi. Setiap konsentrasi memiliki 3 mata kuliah prasyarat dan 7 mata kuliah konsentrasi yang bersifat wajib.

Perancangan & Preprocessing

Gambar 3 merupakan ilustrasi dari rancangan alur sistem yang digunakan untuk implementasi metode *item-based collaborative filtering* pada sistem rekomendasi konsentrasi di STMIK STIKOM Bali. Diawali dengan akuisisi data berupa data nilai mahasiswa. Setelah data diperoleh dilanjutkan dengan tahapan pemrosesan data nilai dengan hanya menggunakan data mahasiswa yang telah memilih konsentrasi dan merepresentasikan data nilai ke dalam bentuk matriks seperti pada Tabel 2.

Data nilai mahasiswa dijadikan landasan pengukuran nilai kemiripan dengan metode *adjusted-cosine similarity* menggunakan Persamaan (1). Hasil pengukuran kemiripan adalah berupa matriks kemiripan yang berikutnya digunakan untuk menentukan *neighborhood* yang relevan untuk setiap item sebagai bahan untuk membuat prediksi nilai terhadap mata kuliah konsentrasi. Komputasi prediksi nilai dihitung dengan metode *weighted sum* menggunakan Persamaan (2). Berdasarkan prediksi nilai untuk mata kuliah konsentrasi, dilakukan penyusunan rekomendasi konsentrasi.



Gambar 3 Desain alur sistem

Rekomendasi disusun berdasarkan indeks prestasi dari prediksi nilai mata kuliah konsentrasi. Konsentrasi yang diprediksi memiliki indeks prestasi konsentrasi lebih besar atau sama dengan 2,75 akan direkomendasikan. Keluaran dari sistem ini berupa urutan rekomendasi konsentrasi yang memiliki prediksi indeks prestasi paling tinggi sampai dengan batas bawah 2,75 dari mata kuliah yang termasuk di dalam mata kuliah konsentrasi. Berdasarkan rekomendasi konsentrasi yang dihasilkan akan dilakukan evaluasi untuk mengetahui efektivitas dan relevansi dari rekomendasi sistem.

Implementasi

Berdasarkan hasil perancangan dapat dilakukan implementasi agar dapat dilakukan evaluasi terhadap kemampuan sistem. Langkah-langkah implementasi terbagi ke dalam tiga tahap. Tahap pertama adalah tahap pengukuran kemiripan *item* yang diukur menggunakan metode

adjusted-cosine similarity. Berdasarkan nilai kemiripan antar *item* diperoleh *item* paling mirip (*neighborhood*) untuk masing-masing *item*. Berdasarkan *neighborhood* dapat dilakukan proses komputasi prediksi di tahap ke-2. Berdasarkan prediksi *nilai* terhadap mata kuliah konsentrasi dapat dilakukan tahap ke-3 yaitu penyusunan rekomendasi konsentrasi.

Pengujian

Pengujian sistem dilakukan untuk mengetahui efektivitas dan relevansi dari rekomendasi yang diberikan oleh sistem. Pengujian model sistem menggunakan metode *k-fold cross validation* untuk membantu proses pembagian data ke dalam data *training* dan data uji. Digunakan nilai $k=10$ untuk membagi data menjadi 10 bagian, di mana sebelum dilakukan pembagian, data diacak terlebih dahulu. Dari 10 bagian data, dilakukan perulangan sebanyak 10 kali untuk melakukan proses *training* dan validasi, Setiap pengulangannya diambil satu segmen dari 10 untuk dijadikan data *testing* dan sisanya menjadi data *training*.

Proses pengukuran kemiripan item dilakukan terhadap data *training* dengan menggunakan metode pengukur kemiripan. *Adjusted-cosine*. Keluaran dari proses pengukur kemiripan adalah matriks kemiripan item yang digunakan untuk menentukan *neighborhood* untuk melakukan komputasi prediksi nilai untuk mata kuliah konsentrasi yang ada pada data testing. Jumlah *neighborhood* yang digunakan pada penelitian ini adalah 20 yang merupakan hasil optimal pada penelitian [10]. Berdasarkan prediksi yang dihasilkan dilakukan proses penyusunan rekomendasi konsentrasi berdasarkan nilai prediksi terhadap mata kuliah konsentrasi.

Suatu konsentrasi akan direkomendasikan jika diprediksi menghasilkan rata-rata nilai berupa indeks prestasi untuk mata kuliah konsentrasi lebih besar atau sama dengan 2,75. Untuk mengetahui efektivitas dan relevansi dari rekomendasi yang diberikan oleh sistem dilakukan pengukuran dengan menggunakan *precision* untuk efektivitas dan *recall* untuk relevansi rekomendasi.

HASIL DAN PEMBAHASAN

Pengujian sistem dilakukan untuk mengetahui efektivitas dan relevansi dari rekomendasi yang diberikan oleh sistem. Efektivitas rekomendasi dihitung dengan menggunakan *precision*. Sementara relevansi rekomendasi dihitung menggunakan *recall*. Sebelum menghitung *precision* dan *recall*, ditentukan objektif relevan yang digunakan pada penelitian ini. Relevan adalah ketika sistem dapat memberikan rekomendasi konsentrasi yang sesuai dengan preferensi mahasiswa.

Pada penelitian ini, sistem memberikan rekomendasi terhadap konsentrasi yang diprediksi menghasilkan nilai berupa indeks prestasi untuk mata kuliah konsentrasi lebih besar atau sama dengan 2,75. Nilai 2,75 digunakan karena nilai tersebut merupakan nilai Indeks Prestasi Akademik minimum untuk memperoleh kelulusan di STIKOM Bali. Berdasarkan ketetapan tersebut maka dibuat tabel *confusion matrix* untuk menghitung nilai *precision* dan *recall*. Tabel *confusion matrix* yang digunakan pada penelitian ini adalah seperti pada Tabel 3.

Tabel 3. *Confusion matrix* yang digunakan

| | Nilai aktual ≥ 2,75 | Nilai aktual < 2,75 |
|-------------------------------|-------------------------------|-------------------------------|
| Direkome- dasikan | <i>True Positive</i> (TP) | <i>False Positive</i> (FP) |
| Tidak direkome- dasikan | <i>False Negative</i> (FN) | <i>True Negative</i> (TN) |

True positive (TP) pada Tabel 3 merupakan kondisi ketika konsentrasi yang direkomendasikan telah ditempuh oleh mahasiswa dan menghasilkan indeks prestasi lebih besar atau sama dengan 2,75 direkomendasikan oleh sistem. *False positive* (FP) merupakan kondisi ketika konsentrasi yang direkomendasikan oleh sistem ditempuh oleh mahasiswa menghasilkan indeks prestasi di bawah 2,75. *False negative* (FN) menggambarkan kondisi ketika konsentrasi yang ditempuh oleh mahasiswa menghasilkan indeks prestasi lebih besar atau sama dengan 2,75

tidak direkomendasikan oleh model sistem. Kondisi terakhir *True Negative* (TN) adalah kondisi ketika konsentrasi yang tidak direkomendasikan oleh sistem ditempuh oleh mahasiswa menghasilkan indeks prestasi di bawah 2,75.

Pengaruh Jumlah *Neighborhood*

Pengujian pertama dilakukan untuk mengevaluasi jumlah *neighborhood* optimal yang digunakan untuk menghasilkan prediksi nilai terbaik. Jumlah *neighborhood* yang digunakan untuk pengujian adalah 3, 5, 10, 15, 20, 25, 30, 35, 40, 45, dan 50. Dimulai dari tiga karena mata kuliah prasyarat yang digunakan di STMIK STIKOM Bali untuk menentukan suatu konsentrasi adalah 3. Pada pengujian ini dilakukan evaluasi jumlah mata kuliah ideal yang berpengaruh untuk pemberian rekomendasi konsentrasi melalui prediksi nilai mata kuliah konsentrasi.

Tabel 4 adalah hasil pengujian pengaruh jumlah *neighborhood* terhadap kualitas prediksi pada data Program Studi Sistem Komputer dan Sistem Informasi. Dari 10 kali pengujian terhadap masing-masing data diperoleh rata-rata galat prediksi paling rendah pada data Program Studi Sistem Komputer adalah 0,4075 yang diperoleh dengan jumlah *neighborhood* 20. Sementara pada data Program Studi Sistem Informasi adalah 0,4283 yang juga diperoleh dengan jumlah *neighborhood* 20. Ukuran *neighborhood* 20 digunakan untuk pengujian berikutnya.

Tabel 4. Hasil pengujian *neighborhood*

| K- neighbors | Sistem Komputer | Sistem Informasi |
|-----------------|--------------------|---------------------|
| 3 | 0,4840 | 0,4919 |
| 5 | 0,4366 | 0,4537 |
| 10 | 0,4116 | 0,4378 |
| 15 | 0,4081 | 0,4285 |
| 20 | 0,4075 | 0,4283 |
| 25 | 0,4094 | 0,4291 |
| 30 | 0,4113 | 0,4293 |
| 35 | 0,4084 | 0,4293 |
| 40 | 0,4109 | 0,4310 |
| 45 | 0,4109 | 0,4303 |
| 50 | 0,4109 | 0,4303 |
| Minimum | 0,4075 | 0,4283 |

Data Sistem Komputer

Hasil dari pengujian pada data Program Studi Sistem Komputer adalah seperti yang ditampilkan pada Tabel 5. Berdasarkan 10 kali pengujian terhadap data Program Studi Sistem Komputer dihasilkan rata-rata nilai *precision* sebesar 0,98. Hasil tersebut bermakna bahwa 98% rekomendasi yang diberikan oleh sistem yang ditempuh oleh mahasiswa dapat menghasilkan indeks prestasi di atas atau sama dengan 2,75.

Tabel 5. Hasil Pengujian Sistem Komputer

| K | TP | FP | FN | TN | Precisi on | Recall |
|------------------|----|----|----|----|-------------|-------------|
| 1 | 21 | 2 | 10 | 2 | 0,91 | 0,68 |
| 2 | 25 | 0 | 8 | 2 | 1,00 | 0,76 |
| 3 | 22 | 1 | 10 | 2 | 0,96 | 0,69 |
| 4 | 30 | 0 | 3 | 2 | 1,00 | 0,91 |
| 5 | 27 | 0 | 6 | 1 | 1,00 | 0,82 |
| 6 | 19 | 0 | 8 | 7 | 1,00 | 0,70 |
| 7 | 24 | 1 | 6 | 3 | 0,96 | 0,80 |
| 8 | 27 | 0 | 6 | 1 | 1,00 | 0,82 |
| 9 | 26 | 0 | 6 | 2 | 1,00 | 0,81 |
| 10 | 19 | 1 | 11 | 3 | 0,95 | 0,63 |
| Rata-rata | | | | | 0,98 | 0,76 |

Dari 10 kali pengujian data Program Studi Sistem komputer rata-rata nilai *recall* yang dihasilkan adalah sebesar 0,76. Nilai tersebut bermakna bahwa 76% dari seluruh mahasiswa yang memperoleh nilai lebih besar atau sama dengan 2,75 pada suatu konsentrasi dapat diprediksi oleh model sistem dengan benar.

Data Sistem Informasi

Hasil dari pengujian *precision* dan *recall* pada data Program Studi Sistem Informasi adalah seperti yang ditampilkan pada Tabel 6. Berdasarkan 10 kali pengujian terhadap data Program Studi Sistem Komputer dihasilkan rata-rata nilai *precision* sebesar 0,98. Hasil tersebut bermakna bahwa 98% rekomendasi yang diberikan oleh sistem yang ditempuh oleh mahasiswa dapat menghasilkan indeks prestasi di atas atau sama dengan 2,75. Dari 10 kali pengujian data Program Studi Sistem komputer rata-rata nilai *recall* yang dihasilkan adalah sebesar 0,73. Nilai

tersebut bermakna bahwa 73% dari seluruh mahasiswa yang memperoleh nilai lebih besar atau sama dengan 2,75 pada suatu konsentrasi dapat diprediksi oleh model sistem dengan benar.

Tabel 6. Hasil Pengujian Sistem Informasi

| K | TP | FP | FN | TN | Precisi on | Recall |
|------------------|----|----|----|----|-------------|-------------|
| 1 | 50 | 1 | 13 | 4 | 0,98 | 0,79 |
| 2 | 50 | 1 | 16 | 1 | 0,98 | 0,76 |
| 3 | 46 | 0 | 17 | 5 | 1,00 | 0,73 |
| 4 | 46 | 0 | 17 | 4 | 1,00 | 0,73 |
| 5 | 52 | 0 | 13 | 2 | 1,00 | 0,80 |
| 6 | 50 | 2 | 11 | 4 | 0,96 | 0,82 |
| 7 | 44 | 1 | 20 | 2 | 0,98 | 0,69 |
| 8 | 43 | 1 | 20 | 3 | 0,98 | 0,68 |
| 9 | 36 | 2 | 28 | 1 | 0,95 | 0,56 |
| 10 | 47 | 2 | 15 | 3 | 0,96 | 0,76 |
| Rata-rata | | | | | 0,98 | 0,73 |

SIMPULAN

Berdasarkan pengujian sistem rekomendasi dengan metode *item-based collaborative filtering* yang telah dilakukan, jumlah *neighborhood* atau mata kuliah yang berpengaruh signifikan untuk membuat prediksi adalah 20 karena dapat menghasilkan galat prediksi paling rendah yang diukur dengan *mean absolute error*. Dengan 20 *neighborhood* untuk membuat prediksi nilai 98% dari rekomendasi konsentrasi yang diberikan oleh sistem terbukti menghasilkan indeks prestasi mata kuliah konsentrasi yang lebih besar atau sama dengan 2,75 pada kedua program studi. Dari seluruh mahasiswa yang memperoleh indeks prestasi untuk mata kuliah konsentrasi lebih besar atau sama dengan 2,75, rekomendasi yang diberikan oleh sistem relevan terhadap 76% mahasiswa pada Program Studi Sistem Komputer dan 73% pada Program Studi Sistem Informasi.

Untuk pengembangan lebih lanjut dapat dilakukan pengembangan kombinasi metode yang dapat meningkatkan kualitas dan relevansi dari sistem rekomendasi. Pengembangan dapat mempertimbangkan penggunaan algoritme genetika untuk memperbaiki hasil pengukuran nilai kemiripan yang diharapkan dapat mening-

katkan kualitas dan relevansi rekomendasi yang dihasilkan.

UCAPAN TERIMAKASIH

Ucapan terima kasih penulis tujukan kepada institusi kami Institut Teknologi dan Bisnis STIKOM Bali sebagai penyandang dana pada penelitian ini melalui surat kontrak penelitian dengan nomor 251/P2M/STIKOM/WDS/V/19.

REFERENSI

- [1] M. I. Dzulhaq and R. Imani, "Sistem Pendukung Keputusan Pemilihan Konsentrasi Jurusan Menggunakan Fuzzy Inference Sistem Metode Mamdani," *Sisfotek Glob.*, vol. 5, no. Sistem Pendukung Keputusan, pp. 75–80, 2015.
- [2] I. Fitrianingrum, "Pengaruh Pemilihan Jurusan dan Motivasi Belajar terhadap Hasil Belajar Siswa Kelas X SMK PGRI 05 Jember Tahun Ajaran 2010/2011," 2014.
- [3] K. B. Anggoro, "Pengaruh Minat Siswa dalam Memilih Program Keahlian Teknik Kendaraan Ringan dan Disiplin Belajar Terhadap Prestasi Belajar Siswa Kelas X Jurusan Teknik Kendaraan Ringan SMK Piri 1 Yogyakarta Tahun Ajaran 2014/2015," Universitas Negeri Yogyakarta, 2015.
- [4] F. Ricci, L. Rokach, and B. Shapira, "Introduction to Recommender Systems Handbook," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Boston, MA: Springer US, 2011, pp. 1–35.
- [5] R. Naughton and X. Lin, "Recommender Systems: Investigation the Impact of Recommendations on User Choices and Behaviours," *ACM RecSys 2010 Work. User-Centric Eval. Recomm. Syst. Their Interfaces*, no. February, pp. 9–13, 2010.
- [6] N. Rosmawarni, N. D. Arianti, and I. Pandini, "Rancang Bangun Aplikasi Sistem Rekomendasi Pencarian Gedung Serbaguna," *SESINDO 2018*, vol. 2018, 2018.
- [7] P. Johansson, "Design and Development of Recommender Dialogue Systems," *Institutionen för datavetenskap*, 2004.
- [8] J. Ben Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative Filtering Recommender Systems," *Adapt. Web Lect. Notes Comput. Sci.*, vol. 4321/2007, pp. 291–324, 2007, doi: DOI: 10.1007/978-3-540-72079-9_9.
- [9] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," *Proc. 10th ...*, vol. 1, pp. 285–295, 2001, doi: 10.1145/371920.372071.
- [10] I. W. Jepriana and S. Hanief, "METODE ITEM-BASED COLLABORATIVE FILTERING UNTUK MODEL SISTEM REKOMENDASI KONSENTRASI PENJURUSAN DI STMIK STIKOM BALI," *J. Teknol. Inf. dan Komput.*, vol. 6, no. 1, 2020.
- [11] A. Wijaya and D. Alfian, "Sistem Rekomendasi Laptop Menggunakan Collaborative Filtering Dan Content-Based Filtering," *J. Comput. Bisnis*, vol. 12, no. 1, pp. 11–27, 2018.
- [12] T. Segaran, *Programming Collective Intelligence: Building Smart Web 2.0 Applications*. 2007.
- [13] A. S. Lampropoulos and G. A. Tsihrantzis, "Machine Learning Paradigms," *Appl. Recomm. Syst. Switz. Springer Intern. Publ*, 2015.
- [14] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation," *Encycl. database Syst.*, pp. 532–538, 2009.
- [15] H. Schütze, C. D. Manning, and P. Raghavan, "Introduction to information retrieval," in *Proceedings of the international communication of association for computing machinery conference*, 2008, vol. 4.