

METODE *ITEM-BASED COLLABORATIVE FILTERING* UNTUK MODEL SISTEM REKOMENDASI KONSENTRASI PENJURUSAN DI STMIK STIKOM BALI

I Wayan Jepriana¹⁾ Shofwan Hanief²⁾

Program Studi Sistem Informasi¹⁾²⁾

Institut Teknologi dan Bisnis STIKOM Bali, Denpasar¹⁾²⁾

jepriana@stikom-bali.ac.id¹⁾ hanief@stikom-bali.ac.id²⁾

ABSTRACT

Concentration selection is a process carried out by STMIK STIKOM Bali students to determine the focus of the specialization of the field to be studied. To take concentration three prerequisite courses are used as a reference whether a student can take a particular concentration. The necessity for students to choose one concentration, while the value of the prerequisite courses is sufficient to choose more than one concentration is often a problem for students. Collaborative Filtering (CF) is a widely used and probably the most common recommendation system technique. In this research, a recommendation system was developed using the item-based collaborative filtering method for the selection of concentrations at STMIK STIKOM Bali. The item-based collaborative filtering method is used because it has the same concept as the current concentration recommendations. Basic courses, prerequisite courses, and concentration courses can be used as items and students who take a course can be used as users in item-based collaborative filtering methods. The evaluation was carried out on the variation of the number of neighborhoods on three similarity measurement methods, namely: cosine similarity, Pearson correlation, and adjusted cosine similarity for student data of Program Studi Sistem Komputer and Program Studi Sistem Informasi. Based on the tests that have been done, it can be concluded that the adjusted cosine similarity measurement method can produce the best predictive quality overall testing based on the generated MAE value. The number of neighborhoods or subjects that have a significant effect on making predictions is 20 because it can produce the lowest prediction errors. After comparing the system models using the similarity measurement method adjusted cosine similarity by using the number of neighborhoods 20 against the predictions produced by the prerequisite courses, it can be concluded that the predictions produced by the 20 neighborhoods are better than the predictions produced by the prerequisite courses.

Keywords: *recommendation system, concentration, item-based collaborative filtering*

ABSTRAK

Pemilihan konsentrasi merupakan suatu proses yang dilakukan oleh mahasiswa STMIK STIKOM Bali untuk menentukan fokus spesialisasi bidang yang ingin dipelajari. Untuk mengambil konsentrasi terdapat tiga mata kuliah prasyarat yang dijadikan acuan apakah seorang mahasiswa bisa mengambil suatu konsentrasi tertentu. Keharusan agar mahasiswa memilih salah satu konsentrasi, sementara nilai dari mata kuliah prasyarat mencukupi untuk memilih lebih dari satu konsentrasi sering menjadi permasalahan bagi mahasiswa. *Collaborative filtering* memanfaatkan informasi rating dari beberapa pengguna untuk memprediksi rating item untuk pengguna tertentu. Pada penelitian ini dilakukan pengembangan sistem rekomendasi dengan metode *item-based collaborative filtering* untuk pemilihan konsentrasi penjurusan di STMIK STIKOM Bali. Metode *item-based collaborative filtering* dipilih karena memiliki konsep yang sama dengan rekomendasi konsentrasi yang berlaku. Mata kuliah dasar, mata kuliah prasyarat dan mata kuliah konsentrasi dapat digunakan sebagai item dan mahasiswa yang mengambil suatu mata kuliah dapat digunakan sebagai user dalam metode *item-based collaborative filtering*. Evaluasi dilakukan terhadap variasi jumlah *neighborhood* pada tiga metode pengukur kemiripan, yaitu: *cosine similarity*, *pearson correlation*, dan *adjusted cosine similarity* untuk data mahasiswa Program Studi Sistem Komputer dan Program Studi Sistem Informasi. Berdasarkan pengujian yang telah dilakukan, dapat

disimpulkan bahwa metode pengukur kemiripan *adjusted cosine similarity* dapat menghasilkan kualitas prediksi yang terbaik secara keseluruhan pengujian berdasarkan galat prediksi yang dihasilkan. Jumlah *neighborhood* atau matakuliah yang berpengaruh signifikan untuk membuat prediksi adalah 20, karena dapat menghasilkan galat prediksi paling rendah. Setelah dilakukan perbandingan antara model sistem yang menggunakan metode pengukur kemiripan *adjusted cosine similarity* dengan menggunakan jumlah *neighborhood* 20 terhadap prediksi yang dihasilkan oleh mata kuliah prasyarat dapat disimpulkan bahwa prediksi yang dihasilkan 20 *neighborhood* lebih baik dari prediksi yang dihasilkan mata kuliah prasyarat.

Kata kunci : *sistem rekomendasi, konsentrasi, item-based collaborative filtering*

PENDAHULUAN

Pemilihan konsentrasi jurusan merupakan suatu proses yang dilakukan oleh mahasiswa STIKOM Bali untuk menentukan fokus spesialisasi bidang yang ingin dipelajari. Setiap konsentrasi menentukan mata kuliah yang harus ditempuh pada semester berikutnya yang spesifik untuk masing-masing konsentrasi. Program Studi S-1 Sistem Komputer di STIKOM Bali memiliki dua konsentrasi jurusan sedangkan Program Studi S-1 Sistem Informasi memiliki empat konsentrasi jurusan. Untuk mengambil setiap konsentrasi terdapat tiga mata kuliah prasyarat yang dijadikan acuan apakah seorang mahasiswa bisa mengambil suatu konsentrasi tertentu. Keharusan agar mahasiswa memilih salah satu konsentrasi, sementara nilai dari mata kuliah prasyarat mencukupi untuk memilih lebih dari satu konsentrasi sering menjadi permasalahan bagi mahasiswa.

Pemilihan jurusan yang tepat di perguruan tinggi dapat membuat suatu perbedaan besar. Mahasiswa perlu menemukan jurusan yang cocok dengan ketertarikan dan kecenderungan kemampuan mereka. Mahasiswa memiliki kemampuan berpikir yang berbeda-beda dan minat yang berbeda-beda pula untuk berbuat sesuatu. Konsentrasi yang tidak sesuai dengan minat mahasiswa bisa menyebabkan kesulitan atau bahkan kegagalan mahasiswa pada suatu matakuliah tertentu yang memberikan dampak pada nilai dari mahasiswa.

Sistem rekomendasi adalah suatu teknik dan perangkat lunak yang dapat memberikan suatu anjuran atau rekomendasi yang sekiranya bermanfaat bagi pengguna dimana anjuran tersebut membantu dalam proses pengambilan keputusan oleh pengguna. Sistem rekomendasi dapat membantu pengguna dalam mengatasi informasi yang meluap dengan memberikan rekomendasi spesifik bagi pengguna dan diharapkan rekomendasi tersebut bisa memenuhi keinginan dan kebutuhan pengguna. *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. *Collaborative filtering* dibagi menjadi dua macam yaitu *user-based collaborative filtering* dan *item-based collaborative filtering*.

Berdasarkan permasalahan pemilihan konsentrasi jurusan yang ada, pada penelitian ini dilakukan pengembangan sistem rekomendasi dengan metode *item-based collaborative filtering* untuk pemilihan konsentrasi jurusan di STMIK STIKOM Bali. Metode *item-based collaborative filtering* dipilih karena sesuai dengan konsep rekomendasi konsentrasi jurusan yang di landasi kemiripan suatu mata kuliah prasyarat dengan mata kuliah konsentrasi. Mata kuliah dasar, mata kuliah prasyarat dan mata kuliah konsentrasi dapat digunakan sebagai item dan mahasiswa yang mengambil suatu mata kuliah dapat digunakan sebagai user dalam metode *item-based collaborative filtering*. Keluaran dari penelitian ini berupa analisa hasil model sistem rekomendasi dengan metode *item-based collaborative filtering* yang diterapkan pada kasus pemilihan konsentrasi jurusan.

TINJAUAN PUSTAKA

Telah dikaji penelitian-penelitian yang berkaitan dengan sistem rekomendasi, metode *collaborative filtering* baik dengan pendekatan *item-based collaborative filtering* maupun dengan pendekatan *user-based collaborative filtering* dan juga metode yang lain yang telah diterapkan pada kasus pemilihan konsentrasi jurusan.

Pada penelitian tentang pengaruh pemilihan jurusan dan motivasi belajar terhadap hasil belajar siswa kelas X SMK PGRI 05 Jember Tahun Ajaran 2010/2011. Metode analisis data menggunakan uji validitas, analisis regresi linear berganda, varian garis regresi, dan uji t. Hasil penelitian menunjukkan bahwa pemilihan jurusan dan motivasi belajar

mempunyai pengaruh secara parsial terhadap hasil belajar siswa [1]. Didukung dengan penelitian [2] dikembangkan sebuah sistem pemberi saran pemilihan konsentrasi di Jurusan Teknik Informatika UII menggunakan *fuzzy inference system* dengan metode Tsukamoto. Pada penelitian ini menghasilkan sebuah sistem yang dapat memberikan saran rekomendasi dalam pemilihan konsentrasi di Jurusan Teknik Informatika FTI UII. Para pengambil keputusan yaitu pengelola jurusan dapat melakukan manipulasi data terhadap aturan pemberian rekomendasi berdasarkan nilai mata kuliah yang mendukung konsentrasi. Mahasiswa dapat melihat rekomendasi dalam bentuk persentase nilai terhadap konsentrasi jurusan yang direkomendasikan berdasarkan nilai matakuliah yang telah ditempuh.

Penelitian lain dari [3] mengembangkan sistem pendukung keputusan pemilihan konsentrasi untuk mahasiswa *international class* di STMIK AMIKOM Yogyakarta dengan menggunakan metode *K-Means Cluster Analysis*. Pada penelitian ini dilakukan pengelompokan mahasiswa kedalam tiga konsentrasi berdasarkan nilainya. Hasil pengelompokan dijadikan landasan untuk pendukung keputusan untuk mahasiswa. Hasil pengelompokan juga dipengaruhi oleh nilai *centroid* awal yang dipakai dan juga jumlah data yang dipakai. Perbedaan pengambilan data pusat *centroid* awal yang dipakai juga mempengaruhi hasil *centroid* akhirnya. Pengembangan sistem pendukung keputusan untuk kasus pemilihan konsentrasi jurusan juga dilakukan oleh [4]. Pada penelitian ini digunakan *fuzzy inference system* dengan metode Mamdani. Penelitian ini menghasilkan sebuah sistem yang dapat membantu siswa yang mendaftar pada STMIK Bina Sarana Global dalam memutuskan jurusan serta konsentrasi. Data yang diolah dalam metode mamdani yaitu data nilai TPA dan nilai Bakat. Data diposes melalui tahap-tahap perhitungan logika *fuzzy* dan memberikan keluaran berupa rekomendasi jurusan untuk siswa.

Sarwar dkk. (2001) melakukan penelitian tentang sistem rekomendasi dengan menggunakan *collaborative filtering*. Dalam penelitian ini dianalisis pendekatan yang berbeda pada algoritme *item-based recommendation*. Percobaan dilakukan dengan menggunakan beberapa teknik yang berbeda untuk menghitung kesamaan antar item dan teknik yang berbeda untuk mendapatkan hasil rekomendasi. Dilakukan eksperimen untuk mengevaluasi hasil penelitian dan dibandingkan dengan pendekatan *k-nearest neighbor*.

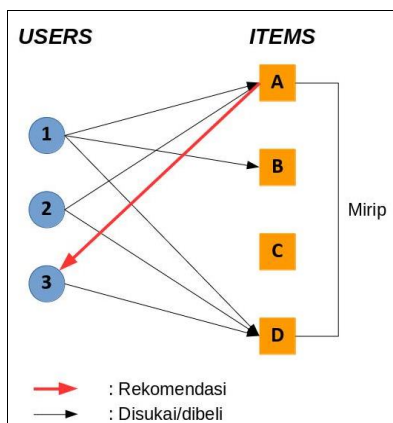
Eksperimen yang telah dilakukan menunjukkan bahwa algoritme *item-based* memberikan kinerja yang lebih baik dari algoritme *user-based*, sementara pada saat yang sama algoritme *item-based* juga memberikan kualitas prediksi yang lebih baik dari pada algoritme *user-based*.

Sistem Rekomendasi

Menurut [6] sistem rekomendasi adalah suatu teknik dan perangkat lunak yang dapat memberikan suatu anjuran atau rekomendasi yang sekiranya bermanfaat bagi pengguna dimana anjuran tersebut membantu dalam proses pengambilan keputusan oleh pengguna seperti memutuskan barang apa yang akan dibeli, musik apa yang akan didengarkan atau berita apa yang akan dibaca. Sistem rekomendasi biasanya fokus pada suatu jenis barang atau *item* dan sistem dirancang sesuai dengan sifat *item* tersebut untuk menghasilkan prediksi rekomendasi yang efektif. Sistem Rekomendasi adalah sistem yang membantu *user* dalam mengatasi informasi yang meluap (*information overload*) dengan memberikan rekomendasi spesifik bagi *user* dan diharapkan rekomendasi tersebut bisa memenuhi keinginan dan kebutuhan *user* [7].

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 (*similarity*) 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 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 [5].



Gambar 1 Ilustrasi *item-based collaborative filtering*

Terdapat dua aspek dalam *item-based collaborative filtering*, yaitu perhitungan kemiripan dan komputasi prediksi. *Item-based collaborative filtering* diilustrasikan seperti pada Gambar 1. Bila dua buah *item* memiliki nilai rating yang mirip diberikan oleh beberapa *user* yang sama maka semakin besar nilai kemiripan antar dua *item* tersebut.

Pengukuran Kemiripan

Salah satu langkah penting dalam metode *item-based collaborative filtering* adalah menghitung kemiripan antara item dan kemudian memilih item yang paling mirip. Metode pengukur kemiripan yang digunakan pada penelitian ini ada tiga, yaitu: *cosine similarity*, *pearson correlation*, dan *adjusted cosine similarity*.

a. Cosine similarity

Pada *cosine similarity* dua item dianggap sebagai dua vektor dalam ruang pengguna m dimensi. Kemiripan di antara keduanya diukur dengan menghitung nilai sudut kosinus antara dua vektor tersebut [5]. Nilai kemiripan item i dengan item j dapat dihitung dengan Persamaan (1)

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

b. Pearson correlation

Dalam kasus ini, kemiripan antara dua item i dan j diukur dengan menghitung korelasi *Pearson-r corr_{i,j}*. Untuk membuat perhitungan korelasi yang akurat, perlu dilakukan isolasi kasus *co-rated* (yaitu, kasus dimana pengguna menilai i dan j). Pengguna yang menilai i dan j dilambangkan dengan U . Untuk menghitung nilai kemiripan dengan *pearson correlation similarity* dilakukan dengan Persamaan (2). Pada Persamaan (2), $R_{u,i}$ menunjukkan

peringkat yang diberikan oleh pengguna u untuk item i , \bar{R}_i adalah nilai rata-rata dari item ke- i [5].

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

c. Adjusted cosine similarity

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 mengura-angi 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 skema ini diberikan oleh Persamaan (3), dimana \bar{R}_u adalah rata-rata rating pengguna ke- u .

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

Komputasi Prediksi

Prediksi rating untuk sebuah item harus didasarkan pada rating yang telah diberikan pengguna untuk *item* yang serupa. Persamaan (4) 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 .

$$\text{pred}(u, i) = \frac{\sum \text{itemSim}(i, j) \cdot r_{uj}}{\sum \text{itemSim}(i, j)} \quad (4)$$

Evaluasi

Untuk menilai performa suatu sistem rekomendasi dibutuhkan suatu metode dan perhitungan yang dapat mengukur tingkat akurasi dari prediksi yang dihasilkan oleh sistem.

d. K-Fold Cross Validation

K-fold cross validation digunakan untuk menghitung akurasi prediksi suatu sistem. Dalam *k-fold cross validation* data dibagi 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 [8].

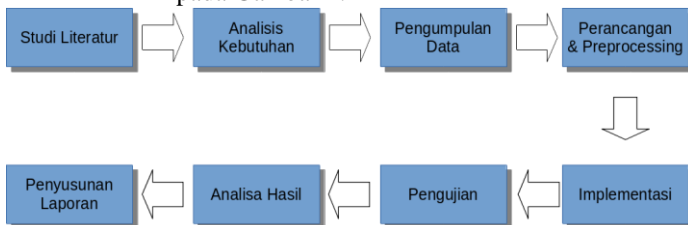
e. Mean Absolute Error

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 (5) sebagai berikut:

$$MAE = \frac{\sum_{i=1}^N |p_i - q_i|}{N} \quad (5)$$

METODOLOGI PENELITIAN

Metode penelitian yang digunakan untuk memodelkan implementasi metode *item-based collaborative filtering* untuk sistem rekomendasi konsentrasi penjurusan di STMIK STIKOM Bali menggunakan tahapan-tahapan dari penelitian yang dilakukan, baik dari tahapan pengumpulan data hingga proses pengujian dan analisa hasil yang telah diperoleh. Tahapan-tahapan yang dilakukan dalam penelitian ini adalah seperti yang digambarkan pada Gambar 2.

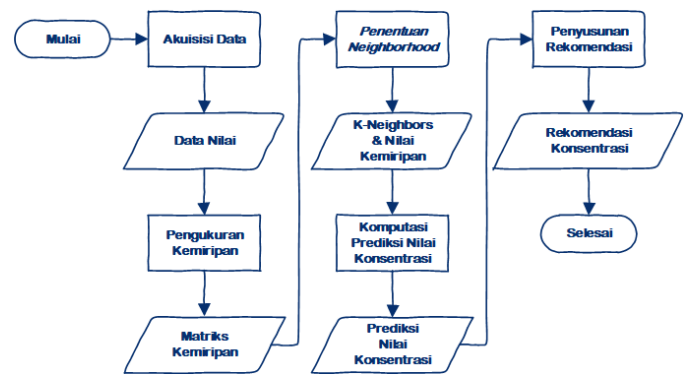


Gambar 2. Metode penelitian

Pada Gambar 2 diilustrasikan bagaimana sistematisa penelitian yang dilakukan. Dimulai dengan studi literatur mengenai literatur yang sekiranya berhubungan dengan topik penelitian. Kemudian dilanjutkan dengan proses analisis kebutuhan, untuk mengetahui kebutuhan apa saja yang diperlukan pada penelitian ini. Berdasarkan analisis kebutuhan, bisa dilanjutkan dengan proses pengumpulan data yang diperlukan untuk penelitian. Setelah data diperoleh dilanjutkan dengan tahapan pengolahan data dan perancangan model sistem. Berdasarkan hasil perancangan dapat dilakukan implementasi model untuk melakukan evaluasi terhadap kemampuan model sistem. Berdasarkan hasil yang diperoleh dari pengujian implementasi model dapat diperoleh hasil untuk penulisan laporan.

Alur Sistem

Gambar 3 merupakan ilustrasi dari usulan model sistem yang digunakan untuk implementasi metode *item-based collaborative filtering* untuk sistem rekomendasi konsentrasi penjurusan di STMIK STIKOM Bali. Diawali dengan akuisisi data berupa data nilai mahasiswa. Data nilai mahasiswa akan dijadikan landasan pengukuran nilai kemiripan dengan pendekatan berbasis *item*. Hasil pengukuran kemiripan adalah berupa matriks kemiripan yang berikutnya digunakan untuk mencari *neighborhood* yang relevan untuk setiap *item* sebagai bahan prediksi nilai. Berdasarkan hasil prediksi nilai, dilakukan penyusunan rekomendasi. Keluaran dari sistem ini berupa rekomendasi konsentrasi penjurusan yang memiliki prediksi nilai paling tinggi dari matakuliah-matakuliah yang termasuk di dalam mata kuliah konsentrasi. Berdasarkan rekomendasi konsentrasi yang dihasilkan dilakukan proses evaluasi untuk mengetahui akurasi dari model sistem.



Gambar 3. Alur sistem

Data Penelitian

Data yang digunakan pada penelitian ini ada dua, yaitu data nilai mahasiswa Program Studi Sistem Komputer dan Program Studi Sistem Informasi. Dari kedua program studi, data mahasiswa angkatan tahun 2015 digunakan karena pada mahasiswa angkatan tahun 2015 rata-rata mahasiswa per program studi yang sudah mengambil konsentrasi diatas 90%. Untuk pengujian model yang diusulkan digunakan data mahasiswa yang telah mengambil konsentrasi.

f. Program Studi Sistem Komputer

Pada data program studi sistem informasi angkatan 2015, terdapat 687 orang mahasiswa. Dari 687 orang mahasiswa, terdiri dari 14 orang mahasiswa belum mengambil konsentrasi dan 673 orang yang sudah mengambil konsentrasi.

Pada program studi sistem komputer terdapat 75 pilihan matakuliah yang terdiri dari 11 mata kuliah pra syarat dan 27 mata kuliah konsentrasi. Setiap konsentrasi memiliki 3 mata kuliah prasyarat dan 7 matakuliah konsentrasi yang bersifat wajib.

g. Program Studi Sistem Informasi

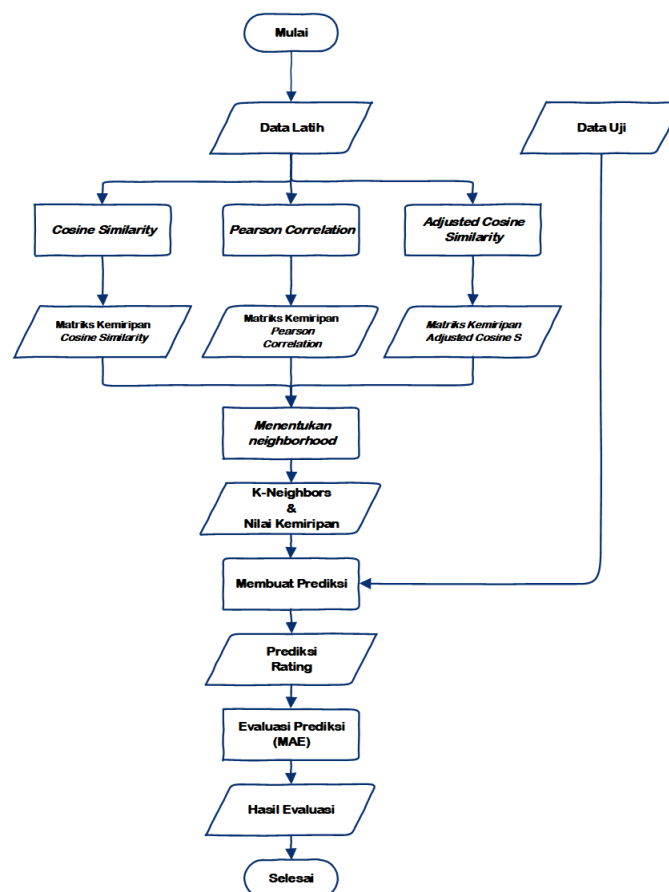
Pada data program studi sistem informasi angkatan 2015, terdapat 687 orang mahasiswa. Dari 687 orang mahasiswa, terdiri dari 14 orang mahasiswa belum mengambil konsentrasi dan 673 orang yang sudah mengambil konsentrasi. Pada program studi sistem komputer terdapat 75 pilihan matakuliah yang terdiri dari 11 mata kuliah pra syarat dan 27 mata kuliah konsentrasi. Setiap konsentrasi memiliki 3 mata kuliah prasyarat dan 7 matakuliah konsentrasi yang bersifat wajib.

Skema Pengujian

Pengujian model sistem dilakukan untuk mengetahui kualitas prediksi yang dihasilkan terhadap nilai matakuliah konsentrasi. Kualitas prediksi dievaluasi dengan mengukur tingkat kesalahan prediksi dengan menggunakan *mean absolute error*. 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, dimana sebelum dilakukan pembagian, data dilakukan proses acak 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*.

Gambar 4 adalah proses yang dilalui oleh setiap perulangan pada metode *k-fold cross validation*. Setelah data di bagi menjadi beberapa segmen dan setiap perulangannya telah ditentukan segmen data yang menjadi data *training* dan data *testing* dilanjutkan dengan proses pengukuran kemiripan *item* terhadap data *training* dengan menggunakan tiga metode pengukur kemiripan. Ketiga metode yang digunakan adalah *cosine similarity*, *pearson correlation*, dan *adjusted cosine*. Keluaran dari masing-masing metode pengukur kemiripan adalah matriks kemiripan *item* yang digunakan untuk menentukan *neighborhood* untuk melakukan komputasi prediksi nilai untuk matakuliah konsentrasi yang ada pada data *testing*. Berdasarkan prediksi yang dihasilkan dilakukan proses pengukuran galat prediksi dengan metode *mean absolute error*. Galat prediksi dari seluruh iterasi digunakan untuk

menentukan rata-rata galat prediksi untuk setiap metode pengukur kemiripan. Selain pengujian untuk perbandingan kualitas prediksi dari 3 metode pengukur kemiripan yang digunakan juga dilakukan pengujian terhadap pengaruh jumlah *neighborhood* yang digunakan untuk komputasi prediksi nilai.



Gambar 4. Alur Pengujian

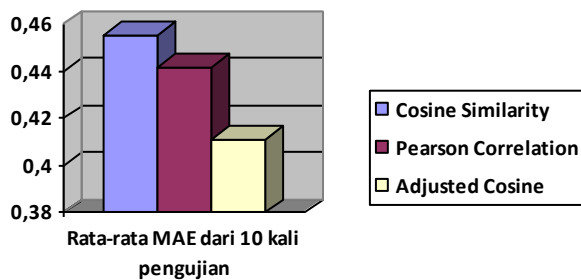
HASIL DAN PEMBAHASAN

Pengujian terhadap model sistem dilakukan dengan tiga tahapan pengujian. Pengujian pertama untuk mencari metode pengukur kemiripan terbaik yang dapat menghasilkan galat prediksi terendah dari kedua data program studi yang digunakan. Pengujian yang kedua adalah variasi jumlah *neighborhood* yang ideal untuk dapat menghasilkan prediksi nilai terhadap matakuliah konsentrasi dengan galat prediksi paling rendah dari ketiga metode pengukur kemiripan. Pengujian terakhir dilakukan perbandingan galat prediksi yang dihasilkan oleh metode pengukur kemiripan terbaik pada pengujian pertama dan jumlah *neighborhood* paling ideal yang di hasilkan pada pngujian

kedua dengan galat prediksi yang dihasilkan oleh mata kuliah prasyarat.

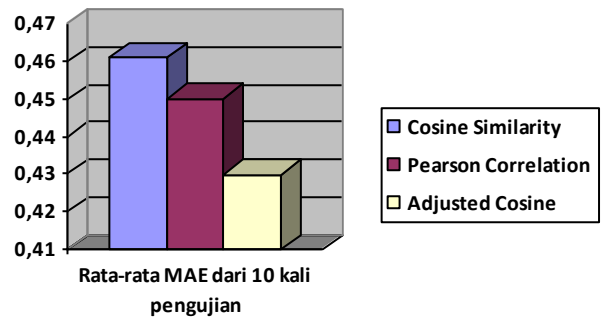
Pengaruh Metode Pengukur Kemiripan

Pengujian pertama dilakukan perbandingan kualitas prediksi yang diukur dengan galat prediksi menggunakan MAE yang dihasilkan oleh ketiga metode pengukur kemiripan yang digunakan. Matriks kemiripan yang dihasilkan oleh metode pengukur kemiripan *cosine similarity*, *pearson correlation*, dan *adjusted cosine* digunakan untuk menghitung prediksi nilai untuk mata kuliah konsentrasi tanpa batasan jumlah *neighborhood*. Masing-masing metode dilakukan pengujian sebanyak 10 kali dan dihitung nilai rata-rata MAE yang dihasilkan. Gambar 5 menunjukkan hasil pengujian terhadap data program studi sistem komputer. Berdasarkan hasil pengujian menunjukkan metode pengukur kemiripan *adjusted cosine similarity* menghasilkan galat prediksi paling rendah dari ketiga metode pengukur kemiripan yang diuji.



Gambar 5. Pengaruh metode pengukur kemiripan terhadap galat prediksi pada data prodi sistem komputer

Hasil pengujian untuk data program studi sistem informasi ditunjukkan pada Gambar 6. Berdasarkan hasil pengujian untuk data program studi sistem informasi juga menunjukkan metode pengukur kemiripan *adjusted cosine similarity* juga menghasilkan galat prediksi yang paling rendah dari ketiga metode pengukur kemiripan yang diuji.



Gambar 6. Pengaruh metode pengukur kemiripan terhadap galat prediksi pada data prodi sistem informasi

Berdasarkan pengujian pengaruh metode pengukur kemiripan yang dilakukan pada kedua data program studi menunjukkan hasil yang sama yaitu metode pengukur kemiripan *adjusted cosine similarity* menghasilkan galat prediksi paling rendah. Hal ini menunjukkan rata-rata nilai dari pengguna yang disesuaikan pada metode *cosine similarity* memberikan keunggulan yang nyata sebagai metode pengukur kemiripan yang ditunjukkan dengan nilai MAE yang dihasilkan secara signifikan lebih rendah.

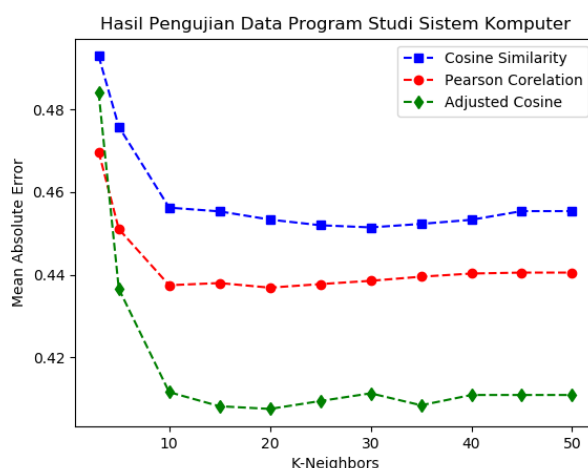
Pengaruh Jumlah Neighborhood

Pengujian yang kedua dilakukan untuk memvalidasi hasil pengujian di tahap pertama. Jika pada tahap pertama dilakukan untuk mencari pengaruh metode pengukur kemiripan tanpa memperhatikan jumlah *neighborhood*, pada pengujian ini dilakukan pengujian terhadap jumlah *neighborhood* yang digunakan untuk menghasilkan prediksi nilai. 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 1 Hasil pengujian pengaruh jumlah *neighborhood* pada data Program Studi Sistem Komputer

K-neighborhoods	Cosine Similarity	Pearson Correlation	Adjusted Cosine
3	0,4929	0,4697	0,4840
5	0,4758	0,4510	0,4366
10	0,4562	0,4374	0,4116
15	0,4553	0,4380	0,4081
20	0,4533	0,4368	0,4075
25	0,4520	0,4377	0,4094
30	0,4514	0,4385	0,4113
35	0,4523	0,4395	0,4084
40	0,4533	0,4403	0,4109
45	0,4554	0,4405	0,4109
50	0,4554	0,4405	0,4109
Minimum	0,4514	0,4368	0,4075

Tabel 1 adalah hasil pengujian pengaruh jumlah *neighborhood* terhadap kualitas prediksi dari masing-masing metode pengukur kemiripan pada data Program Studi Sistem Komputer. Dari 10 kali pengujian terhadap masing-masing metode pengukur kemiripan diperoleh rata-rata galat prediksi paling rendah pada metode pengukur kemiripan *cosine similarity* diperoleh pada jumlah *neighborhood* 30 dengan MAE sebesar 0,4514. Galat prediksi paling rendah untuk metode *pearson correlation* diperoleh pada jumlah *neighborhood* 20 dengan MAE sebesar 0,4368. Sedangkan galat prediksi terendah yang diperoleh metode *adjusted cosine similarity* adalah 0,4075 yang diperoleh dengan jumlah *neighborhood* 20.



Gambar 7. Sensitivitas jumlah *neighborhood* pada data Program Studi Sistem Komputer

Tabel 1 direpresentasikan dalam bentuk grafik ditunjukkan pada Gambar 7. Secara garis besar, metode pengukur kemiripan *adjusted cosine similarity* jauh mengungguli dua metode pengukur kemiripan lainnya. Hanya pada jumlah *neighborhood* 3 metode *pearson correlation* lebih unggul. Nilai MAE terendah secara keseluruhan dari ketiga metode pengukur kemiripan dan variasi jumlah *neighborhood* dicapai oleh metode pengukur kemiripan *adjusted cosine similarity* dengan jumlah *neighborhood* 20.

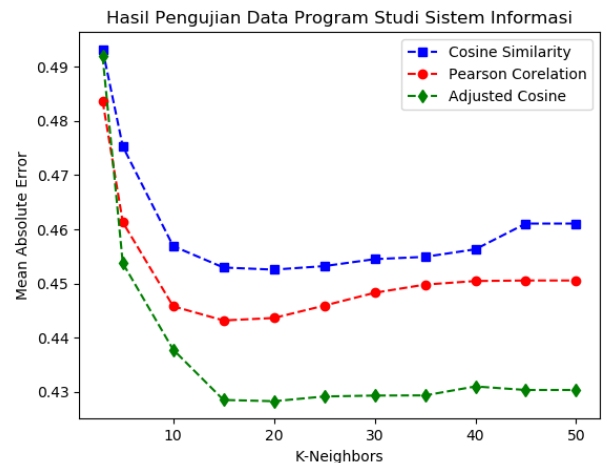
Tabel 2 Hasil pengujian pengaruh jumlah *neighborhood* pada data Program Studi Sistem Informasi

K-neighbors	Cosine Similarity	Pearson Correlation	Adjusted Cosine
3	0,4932	0,4836	0,4919
5	0,4753	0,4613	0,4537
10	0,4569	0,4458	0,4378
15	0,4529	0,4432	0,4285
20	0,4526	0,4436	0,4283
25	0,4532	0,4459	0,4291
30	0,4545	0,4483	0,4293
35	0,4549	0,4498	0,4293
40	0,4563	0,4505	0,4310
45	0,4611	0,4505	0,4303
50	0,4611	0,4505	0,4303
Minimum	0,4526	0,4432	0,4283

Tabel 2 adalah hasil pengujian pengaruh jumlah *neighborhood* terhadap kualitas prediksi dari masing-masing metode pengukur kemiripan pada data Program Studi Sistem Informasi. Dari 10 kali pengujian terhadap masing-masing metode pengukur kemiripan diperoleh rata-rata galat prediksi paling rendah pada metode pengukur kemiripan *cosine similarity* diperoleh pada jumlah *neighborhood* 20 dengan MAE sebesar 0,4526. Galat prediksi paling rendah untuk metode *pearson correlation* diperoleh pada jumlah *neighborhood* 15 dengan MAE sebesar 0,4432. Sedangkan galat prediksi terendah yang diperoleh metode *adjusted cosine similarity* adalah 0,4283 yang diperoleh dengan jumlah *neighborhood* 20.

Hasil pengujian pada Tabel 2 direpresentasikan dalam bentuk grafik ditunjukkan pada Gambar 8. Hasil yang ditunjukkan secara garis besar sama dengan hasil pengujian untuk data Program Studi Sistem Komputer pada Gambar 7. Metode pengukur kemiripan *adjusted cosine similarity* jauh mengungguli dua metode pengukur kemiripan lainnya. Hanya pada jumlah *neighborhood* 3 metode *pearson correlation* lebih unggul. Nilai MAE terendah secara keseluruhan dari ketiga metode pengukur kemiripan dan variasi jumlah *neighborhood* juga dicapai oleh metode pengukur kemiripan *adjusted cosine similarity* dengan jumlah *neighborhood* 20.

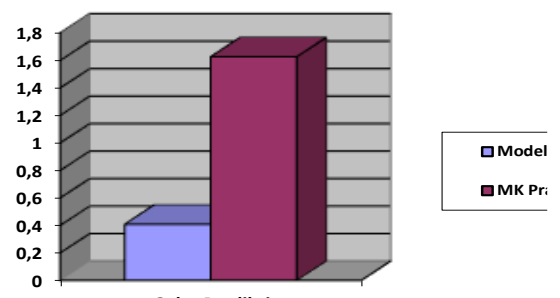
Dari pengujian sensitivitas jumlah *neighborhood* terhadap kedua data yang telah dilakukan. Dapat disimpulkan jumlah mata kuliah yang memberikan pengaruh terbaik terhadap prediksi nilai mata kuliah konsentrasi adalah 20. Sehingga pada pada pengujian berikutnya akan dilakukan pengujian terhadap kualitas prediksi yang dihasilkan model sistem rekomendasi dengan parameter yang telah diperoleh dari hasil pengujian sebelumnya dengan kualitas prediksi yang dihasilkan mata kuliah prasyarat.



Gambar 8. Sensitivitas jumlah neighborhood pada data Program Studi Sistem Informasi

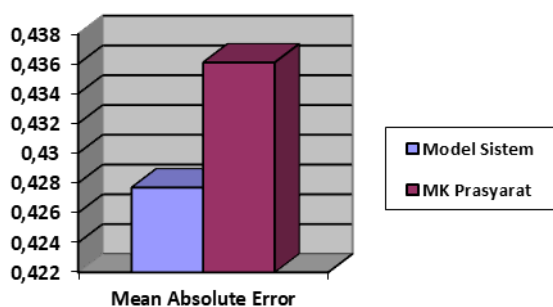
Pengujian Kualitas Prediksi

Setelah dilakukan penentuan metode dan parameter optimal untuk model sistem, selanjutnya akan dilakukan pengujian kualitas prediksi yang dihasilkan oleh model sistem dengan kualitas prediksi yang dihasilkan oleh mata kuliah prasyarat. Hasil pengujian untuk data Program Studi Sistem Komputer digambarkan seperti pada Gambar 9. Galat prediksi yang dihasilkan oleh model sistem secara signifikan lebih rendah dari pada galat prediksi yang diperoleh dari mata kuliah prasyarat. Hal itu menunjukkan bahwa prediksi yang dihasilkan model sistem secara signifikan lebih baik dari pada prediksi yang di hasilkan mata kuliah prasyarat.



Gambar 9. Perbandingan kualitas prediksi untuk data Program Studi Sistem Komputer

Hasil serupa juga dihasilkan pada pengujian terhadap data Program Studi Sistem Informasi. Model sistem dapat menghasilkan kualitas prediksi yang lebih baik dari pada prediksi yang dihasilkan oleh mata kuliah prasyarat. Sehingga dapat disimpulkan bahwa model sistem rekomendasi dengan metode *item-based collaborative filtering* yang menggunakan metode pengukur kemiripan *adjusted cosine similarity* mampu menghasilkan kualitas prediksi yang lebih baik dari pada prediksi dengan mata kuliah prasyarat pada kedua data program studi yang diujikan.



Gambar 10. Perbandingan kualitas prediksi untuk data Program Studi Sistem Informasi

SIMPULAN

Berdasarkan pengujian model sistem rekomendasi dengan metode *item-based collaborative filtering* yang telah dilakukan, dapat disimpulkan bahwa metode pengukur kemiripan *adjusted cosine similarity* dapat menghasilkan kualitas prediksi yang terbaik secara keseluruhan pengujian berdasarkan galat prediksi yang dihasilkan. Jumlah *neighborhood* atau matakuliah yang berpengaruh signifikan untuk membuat prediksi adalah 20 karena dapat menghasilkan galat prediksi paling rendah. Setelah dilakukan perbandingan antara model sistem yang menggunakan metode pengukur kemiripan *adjusted cosine similarity* dengan menggunakan jumlah *neighborhood* 20 terhadap prediksi yang dihasilkan oleh mata kuliah prasyarat dapat disimpulkan bahwa prediksi yang dihasilkan 20 *neighborhood* lebih baik dari prediksi yang dihasilkan mata kuliah prasyarat.

DAFTAR PUSTAKA

- [1] I. Fitrianingrum, "Pengaruh Pemilihan Jurusan dan Motivasi Belajar terhadap Hasil Belajar Siswa Kelas X SMK PGRI 05 Jember Tahun Ajaran 2010/2011," 2014.
- [2] A. Z. Rakhman, H. N. Wulandari, G.

Maheswara, and S. Kusumadewi, "Fuzzy Inference System Dengan Metode Tsukamoto Sebagai Pemberi Saran Pemilihan Konsentrasi (Studi Kasus : Jurusan Teknik Informatika Uii)," *Semin. Nas. Apl. Teknol. Inf. 2012 (SNATI 2012)*, vol. 2012, no. Snati, pp. 15–16, 2012.

- [3] J. Aranda and W. A. G. Natasya, "Penerapan Metode K-Means Cluster Analysis Pada Sistem Pendukung Keputusan Pemilihan Konsentrasi Untuk Mahasiswa International Class Stmik Amikom Yogyakarta," *Semnasteknomedia Online*, vol. 4, no. 1, pp. 4-2-1, 2016.
- [4] M. I. Dzulhaq and R. Imani, "Sistem Pendukung Keputusan Pemilihan Konsentrasi Jurusan Menggunakan Fuzzy Inference System Metode Mamdani," *Sisfotek Glob.*, vol. 5, no. Sistem Pendukung Keputusan, pp. 75–80, 2015.
- [5] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," *Proc. 10th ...*, vol. 1, pp. 285–295, 2001.
- [6] 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.
- [7] 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.
- [8] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation," *Encycl. database Syst.*, pp. 532–538, 2009.