

## Collaborative Filtering untuk memprediksi score Ujian Akhir Siswa di sistem pembelajaran Elektronik

Eko Travada Suprpto Putro  
Fakultas Ilmu Komputer  
Universitas Nasional PASIM  
[travada@pasim.ac.id](mailto:travada@pasim.ac.id)  
[ekotravada@gmail.com](mailto:ekotravada@gmail.com)

### Abstrak

Dalam suatu kegiatan perkuliahan yang menjadi meta data untuk mengukur kemampuan siswa dalam suatu proses pembelajaran adalah komponen penilaian seperti kehadiran, Tugas , UTS dan UAS. Kelengkapan nilai menjadi hal yang sangat penting untuk dapat mengukur kemampuan siswa. Untuk itu nilai yang kosong harus diisi dengan mempertimbangkan nilai siswa tersebut dan dilihat bobot kedekatannya dengan siswa lain untuk dapat memprediksi nilai yang kosong. Dalam paper ini akan dibahas cara memprediksi nilai yang kosong dengan Teknik collaborative filtering yang juga merupakan kebaruan dalam teknik melengkapi penilaian siswa.

Kata Kunci : Collaborative Filtering, Relasi Pearson, Prediksi

### Abstract

In a lecture activity that becomes meta data to measure students' abilities in a learning process is an assessment component such as attendance, assignments, midterm and final exam. The completeness of values becomes very important to be able to measure the ability of students. For that the blank value must be filled by considering the value of the student and the weight of his close relationship with other students to be able to predict the empty value. In this paper, we will discuss how to predict empty scores with collaborative filtering techniques which are also new in the technique of completing student assessments.

Key Word : Collaborative Filtering, Person Correlation, Prediction

### 1. Pendahuluan

Teknik collaborative filtering merupakan suatu cara dalam recommender system untuk dapat merekomendasikan item berdasarkan kedekatan terhadap user tersebut ke user lain dengan mengukur bobot kedekatannya dan kemudian dirangking dari bobot terdekat hingga bobot terjauh [1]. Implementasi Recommender system dengan Teknik collaborative filtering dimulai dari penelitian oleh bobadila [2]. Dalam penelitian ini bobadila melihat kedekatan antar user dengan

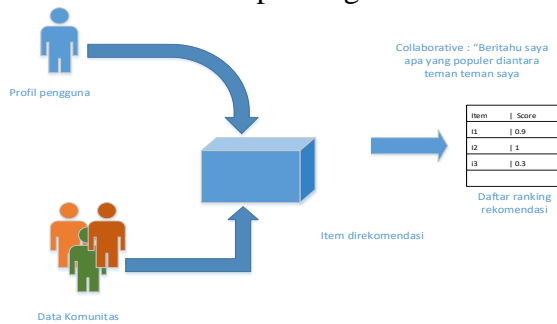
mengukur kedekatan berdasar bobot setiap user dengan melihat nilai bobot tes, bobot materi pelajaran dan bobot dari pengajar setiap siswa. Hal yang menjadi masalah dalam pengukuran kedekatan antar setiap user dalam collaborative filtering adalah tidak lengkapnya data dari setiap user [3] dan berakibat menurunnya kualitas dari suatu proses rekomendasi. Untuk memperbaiki kondisi ini maka perlu dilakukan prediksi data yang kosong agar dapat memperbaiki akurasi dari collaborative filtering. Ketidaklengkapan data dalam sistem

pembelajaran umumnya terjadi dalam latihan pembelajaran karena siswa umumnya cenderung lebih suka membaca materi di seluruh subbab dengan mengabaikan latihan dari setiap sub bab tersebut [4][5]. Teknik memprediksi dengan Teknik collaborative filtering diusulkan sebagai cara untuk memprediksi nilai yang kosong dan merupakan suatu kebaruan untuk diimplementasikan di sistem pembelajaran elektronik.

**2. Teori Dasar**

**2.1. Collaborative Filtering**

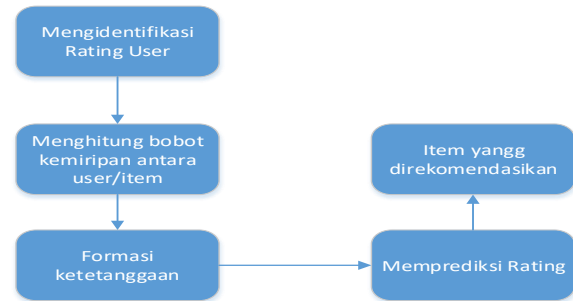
Gagasan utama dari pendekatan rekomendasi kolaboratif adalah untuk mengeksplorasi informasi tentang perilaku masa lalu atau pendapat dari komunitas pengguna yang ada untuk memprediksi item mana yang mungkin akan disukai atau diminati oleh pengguna sistem saat ini. Tampak di gambar 1



Gambar 1 Collaborative Filtering [6]

Pendekatan kolaboratif murni mengambil matriks peringkat item pengguna yang diberikan sebagai hanya input dan biasanya menghasilkan jenis keluaran berikut: (a) prediksi (numerik) yang menunjukkan sejauh mana pengguna akan menyukai atau tidak menyukai item tertentu dan (b) daftar item yang direkomendasikan. Daftar top-N seperti itu seharusnya, tentu saja, tidak mengandung item yang sudah dibeli oleh pengguna saat

ini. Arsitektur Collaborative Filtering secara umum tampak di gambar 2



Gambar 2 Arsitektur Collaborative Filtering [7]

**2.2. User Based Nearest Neighbor (NN) Recommendation**

User Based Nearest Neighbor Recommendation adalah metoda yang digunakan dalam collaborative filtering dalam menentukan user lain yang memiliki kedekatan terbaik ke user yang sedang aktif. Gagasan utama User Based Nearest Neighbor Recommendation adalah sebagai berikut: diberi basis data peringkat dan ID dari pengguna saat ini (aktif) sebagai input, identifikasi pengguna lain (kadang-kadang disebut sebagai pengguna sebaya atau tetangga terdekat) yang memiliki preferensi serupa dengan yang dimiliki pengguna aktif di masa lalu. Gunakan  $U = \{u_1, \dots, u_n\}$  untuk menunjukkan set pengguna,  $P = \{p_1, \dots, p_m\}$  untuk  $i \in 1 \dots n, j \in 1 \dots m$ . Nilai-nilai peringkat yang mungkin ditentukan pada numerik seperangkat produk (item), dan  $R$  sebagai matriks  $n \times m$  dari peringkat  $r_{i,j}$  dengan skala dari 1 (sangat tidak suka) hingga 5 (sangat suka). Jika pengguna  $i$  belum memberi nilai item  $j$ , entri matriks yang sesuai,  $r_{i,j}$  tetap kosong. Untuk penentuan yang serupa digunakan Koefisien korelasi Pearson dengan  $sim(a, b)$  dari user  $a$  dan  $b$  di matrik  $R$  dalam formula (1)

$$sim(a, b) = \frac{\sum_{p \in P} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P} (r_{b,p} - \bar{r}_b)^2}} \quad (1)$$

Korelasi pearson akan menghasilkan nilai -1 hingga 1. Nilai dari -1 hingga sebelum 0 menunjukkan bahwa antara user tidak memiliki keserupaan sedangkan nilai 0 hingga menunjukkan bahwa user adalah keserupaan sesuai bobot yang dihasilkan. Bila menghasilkan nilai 1 maka antara user serupa 100 % [8].

**2.3. Prediksi Nilai**

Dalam proses mencari bobot antar user dibutuhkan data yang lengkap agar proses NN akurat. Dalam Implementasi sering kali data tidak lengkap karena dalam sistem pembelajaran elektronik memungkinkan user dalam membaca materi tidak berurut dan tidak selalu mengerjakan sub tugas yang tersedia tapi memilih melanjutkan membaca materi berikutnya. Kondisi data yang tidak lengkap ini tentunya akan menyulitkan untuk memprediksi kemampuan siswa. Agar dapat memprediksi kemampuan siswa maka data harus lengkap dan untuk melengkapi data ini maka nilai yang kosong harus diprediksi. Teknik untuk memprediksi data yang kosong salahh satu pendekatan yang bisa digunakan yaitu menggunakan Teknik collaborative filtering. Teknik collaborative filtering untuk memprediksi data dilakukan oleh cs-lee [9] untuk memprediksi model dengan melakukan data mining. Kemudian penelitian yang dilakukan oleh x.su [10] yang melakukan studi literatur Teknik prediksi dengan collaborative filtering. Untuk melakukan prediksi dapat digunakan persamaan (2) yang merupakan Teknik NN dengan nilai a adalah referensi dan b adalah perbandingan terdekat yang didapat dari proses korelasi Pearson.

$$pred(a, b) = \bar{r}_a + \frac{\sum_{b \in N} sim(a,b) * (r_{b,p} - \bar{r}_b)}{\sum_{b \in N} sim(a,b)} \quad (2)$$

**2.4. Mengukur Akurasi**

Untuk mengetahui akurasi dari prediksi data yang telah dihitung perlu dihitung akurasinya terhadap data sebenarnya. Untuk mengukur akurasi prediksi dihitung dengan formula (3)

MAE (Mean Absolute Error).

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

$p_i$  adalah nilai prediksi dan  $r_i$  adalah nilai sebenarnya dan N adalah jumlah data.

**3. Methodology Penelitian**

Penelitian untuk memprediksi nilai menggunakan methodology DSRM [11] dengan penyederhanaan langkah seperti Berikut :

1. Menentukan identifikasi masalah
2. Membuat hipotesis
3. Menguji Hipotesis
4. Melakukan analisa dari hasil pengujian
5. Membuat kesimpulan

**3.1. Metode Penelitian**

Metode untuk memprediksi nilai yang kosong dengan terlebih dahulu mengukur bobot kedekatan antara user yang nilainya tidak lengkap dengan user lain dengan mengukur bobot hubungan dengan korelasi pearson di data yang ada. Nilai bobot user yang terdekat yaitu yang mendekati 1 yang digunakan untuk memprediksi nilai yang masih kosong. Jumlah user yang dibutuhkan untuk memprediksi nilai agar akurat yang menjadi fokus penelitian di paper ini



Gambar 3 Metode Penelitian prediksi nilai

### 3.2. Identifikasi Masalah

Identifikasi masalah dari penelitian ini adalah

1. Bagaimana metode memprediksi data yang tidak lengkap dengan Teknik collaborative filtering
2. Berapa jumlah user yang dibutuhkan dalam collaborative filtering agar prediksi lebih akurat.

### 3.3. Hipotesis

Metode collaborative filtering dapat memprediksi data nilai yang kosong dengan mengambil jumlah user minimal 50 % dari user yang memiliki bobot lebih besar dari 0.

### 3.1. Menguji Hipotesis

Pengujian hypothesis dilakukan dengan langkah seperti Berikut :

1. Mengkonsep disain Eksperimen
2. Eksperimen
3. Merekam hasil eksperimen kemudian melakukan analisis

#### 3.1.1. Konsep Disain Eksperimen

#### Perangkat Lunak dan Data yang digunakan.

Dalam Eksperimen ini perangkat lunak yang digunakan yaitu :

1. Bahasa pemrograman Python
2. Perangkat Lunak Microsoft Excell

Data yang digunakan untuk eksperimen :

1. Data nilai siswa mahasiswa prodi Teknik informatika
2. 2 mata kuliah di prodi Teknik Informatika

### 3.2. Langkah Eksperimen

1. Menetapkan materi kuliah yang akan digunakan untuk memprediksi nilai
2. Mengambil 70 % dan 80 % nilai mata kuliah kemudian ditutup beberapa nilai sebagai data training

diasumsikan sebagai data nilai yang tidak lengkap

3. Data training diukur akurasi dan diambil akurasi yang tertinggi
4. Akurasi yang tertinggi diuji cobakan kembali ke data testing yaitu sebesar 30 % dan 20 % data yang tersisa
5. Dilakukan analisis dari dua model data dipilih yang terbesar akurasi.

#### 3.2.1. Model eksperimen

Eksperimen dalam melakukan prediksi data yang tidak lengkap dalam paper ini dibentuk dengan contoh model seperti tampak di gambar 4

	Tugas 1	Tugas 2	Tugas 3	Tugas 4	Tugas 5
Rudi	80	70	60	50	
User2	80	75	65	60	80
User3	70	50		90	70
User4	75	60	70	60	80
User5	80	80	90	50	90

Gambar 4 Model tabel prediksi data

Tampak di gambar 4 setiap user memiliki nilai tugas dan ada 2 siswa yang memiliki nilai yang tidak lengkap yaitu Rudi dan User 3. Untuk melengkapi nilai Rudi maka Rudi dibandingkan dengan user 2, 4 dan 5. Formula yang digunakan untuk menghitung bobot yaitu person correlation. Bobot yang didapat dengan person Correlation digunakan untuk memilih user yang akan digunakan untuk memprediksi nilai yang belum lengkap dengan formula (2). User yang dipilih adalah user yang memiliki bobot range  $[0..1]$ .

### 4. Tahapan Eksperimen

Dalam eksperimen data yang digunakan yaitu data nilai mahasiswa di 1 mata kuliah (gambar5). Data mata kuliah di gambar 5 tersebut akan diukur bobot relasinya antara seorang siswa dengan siswa yang lain.

	Absensi	Tugas	UTS	UAS
Deden	100	90	85	UNK
Angga	95	90	90	65
Cecep	90	90	90	70
Deni	100	70	80	70
Kusmayadi	85	90	75	60
Lia Azhar	40	0	70	0
Netralius	100	90	70	50
Sivinus zebua	90	90	70	60
Berkat Rahmat	100	60	75	60
Cahyadi	95	90	90	50
Cepi	95	60	75	50
Erlangga	85	60	85	75
Etalisan	100	90	90	50

Gambar 5 Data siswa yang akan diprediksi

Nilai yang akan diprediksi adalah nilai Deden untuk kolom UAS berisi UNK yang artinya belum diketahui. Langkah pertama menghitung bobot dengan persamaan (1) dengan menggunakan data di kolom Absen hingga UTS. Hasil yang didapat tampak di gambar 6.

Similarity(Target,userlain)	Bobot
sim(Deden,Angga)	0.9449
sim(Deden,Cecep)	#DIV/0!
sim(Deden,Deni)	0.7857
sim(Deden,Kusmayadi)	0.5000
sim(Deden,Lia Azhar)	-0.2485
sim(Deden,Netralius)	0.9286
sim(Deden,Sivinus zebua)	0.7559
sim(Deden,Berkat Rahmat)	0.7559
sim(Deden,Cahyadi)	0.9449
sim(Deden,Cepi)	0.7146
sim(Deden,Erlangga)	0.1890
sim(Deden,Etalisan)	0.9449

Gambar 6 pengukuran bobot kedekatan

Dari hasil pengukuran bobot di gambar 6 dipilih beberapa pasang siswa yang bernilai

bobot mendekati 1 yaitu Angga, Etalisan, Cahyadi dan Netralius. Dengan menggunakan persamaan (2) dengan menggunakan beberapa model kombinasi didapat prediksi skor UAS di gambar 7.

Pasangan prediksi	Prediksi Skor Deden
Pred(Angga,Cahyadi)	58
Pred(Angga,Etalisan)	57
Pred(Angga,Netralius)	60
Pred(Cahyadi,Etalisan)	49

Gambar 7 Prediksi Skor dengan kedua bobot dekat 1

Kemudiann diujicobakan dengan siswa yang salah satu memiliki bobot mendekati 1 tampak di gambar 8

Pasangan Prediksi	Prediksi Skor Deden
Pred(Kusmayadi,Etalisan)	47
Pred(Berkat Rahmat,Etalisan)	46
Pred(Kusmayadi,Cahyadi)	48
Pred(Cahyadi,Etalisan)	49

Gambar 8 Prediksi skor dengan salah satu bobot dekat 1

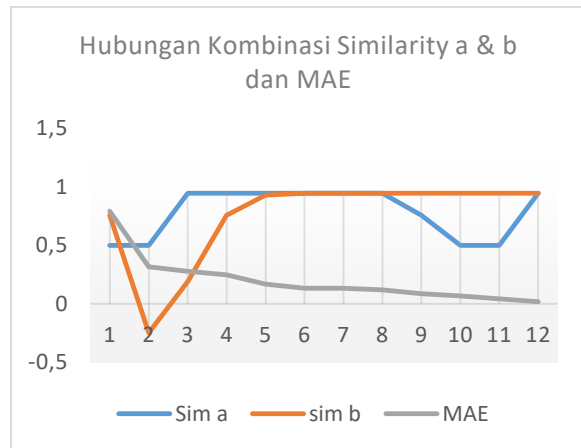
Skor Deden berdasarkan hasil pengujian sebenarnya mendapatkan nilai 50, maka nilai MAE seperti dipersamaan (3) tampak di gambar 9.

Prediksi Pasangan	MAE
Pred(Angga,Cahyadi)	13%
Pred(Angga,Etalisan)	12%
Pred(Angga,Cahyadi)	13%
Pred(Angga,Netralius)	17%
Pred(Cahyadi,Etalisan)	2%
Pred(Kusmayadi,Berkat Rahmat)	21%
Pred(Kusmayadi,Etalisan)	7%
Pred(Berkat Rahmat,Etalisan)	8%
Pred(Kusmayadi,Cahyadi)	4%

Gambar 9 Nilai MAE dari setiap pasangan

Nilai MAE yang akurat adalah yang mendekati nilai 0. Terlihat dari gambar 9 bahwa bila dipilih user yang salah satu

memiliki kemiripan mendekati 1 maka MAE mendekati 0. Secara grafik hubungan antara nilai bobot dan MAE terlihat di gambar 10 bahwa bobot kedua pasangan mendekati 1 maka MAE menuju ke 0.



Gambar 10 Hubungan MAE dan kombinasi similarity

## 5. Kesimpulan

1. Dari percobaan menunjukkan bahwa Collaborative Filtering dapat digunakan untuk memprediksi data nilai yang tidak lengkap
2. Nilai bobot antara user target dan user lain mendekati 1 maka MAE menuju 0
3. Jika salah satu pasangan user memiliki bobot mendekati 1 maka ada kecenderungan MAE 0

## 5. Penelitian ke Depan

Penelitian berikutnya yang bisa dilakukan adalah menentukan presisi dan recall dari prediksi yang dilakukan dengan mengkambil data dari beberapa dosen agar dapat terlihat konsistensi dari prediksi .

## Referensi

- [1] F. O. Isinkaye, "Recommendation systems : Principles , methods and evaluation," pp. 261–273, 2015.
- [2] J. Bobadilla, F. Serradilla, and A. Hernando, "Collaborative filtering adapted to recommender systems of e-learning," *Knowledge-Based Syst.*, vol. 22, no. 4, pp. 261–265, 2009.
- [3] M. Nilashi, O. Ibrahim, and K. Bagherifard, "A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques," *Expert Syst. Appl.*, vol. 92, pp. 507–520, 2018.
- [4] P. Kirschner, J. Sweller, and R. Clark, "Why Unguided Learning Does Not Work," *Educ. Psychol.*, vol. 41, no. 2, pp. 75–86, 2006.
- [5] J. Sweller, "Cognitive load during problem solving: Effects on learning," *Cogn. Sci.*, vol. 12, no. 2, pp. 257–285, 1988.
- [6] D. Jannach and T. U. Dortmund, *Recommender Systems An introduction Recommender Systems*. 2014.
- [7] E. Çano and M. Morisio, "Hybrid recommender systems: A systematic literature review," *Intell. Data Anal.*, vol. 21, no. 6, pp. 1487–1524, 2017.
- [8] E. Tnay, A. E. A. Othman, H. C. Siong, and S. L. O. Lim, "The Influences of Job Satisfaction and Organizational Commitment on Turnover Intention," *Procedia - Soc. Behav. Sci.*, vol. 97, pp. 201–208, 2013.
- [9] C. S. Lee, "Diagnostic, predictive and compositional modeling with data mining in integrated learning environments," *Comput. Educ.*, vol. 49, no. 3, pp. 562–580, 2007.
- [10] X. Su and T. M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques," *Adv. Artif. Intell.*, vol. 2009, no. Section 3, pp. 1–19, 2009.

- [11] K. Peffers *et al.*, “The Design Science Research Process: A Model for Producing and Presenting Information Systems Research,” *Proc. First Int. Conf. Des. Res. Inf. Syst. Technol.*, no. May 2014, pp. 83–106, 2006.
- [12] H. H. Hoos, “Machine Learning – Opportunities and Limitations,” 2017.