

Vol.1 No.1 | Februari 2019 | p-ISSN 2655-867X
e-ISSN 2655-8661

SisInfo

Jurnal Sistem Informasi dan Informatika



Penerbit :

**FAKULTAS TEKNOLOGI DAN INFORMATIKA
UNIVERSITAS INFORMATIKA DAN BISNIS INDONESIA**

DAFTAR ISI

Analisis Faktor-Faktor Yang Mempengaruhi Penggunaan <i>E-Marketplace</i> Shopee Graha Prakarsa	1 – 11
Sistem Informasi Tempat Kost Berbasis Webservice Di Sekitar Universitas Islam Nusantara Menggunakan Metode SAW (<i>Simple Additive Weighting</i>) Hendriyana, Ripal Maulana	12 – 18
Pengembangan Aplikasi Lagu Daerah Dan Nasional Berbasis Android Ivan Michael Siregar	19 – 24
Studi Komparasi Algoritma Similaritas Pada Prediksi Rating Berbasis Item Pada <i>Collaborative Filtering</i> Studi Kasus Pada <i>Data Review</i> Restoran Mochamad Iqbal Ardimansyah	25 – 29
Pengembangan Media Pembelajaran Limit Fungsi Berbasis Multimedia Untuk Sekolah Menengah Atas Kelas XI Marwondo1, R. Yadi Rakhman A, Wilner Saut Lamhot	30 – 40
Aplikasi Pengolahan Data Perusahaan Berbasis Web Dengan Menggunakan <i>Framework Codeigniter</i> (Studi Kasus Cv. Prima Nusa) Tarsinah Sumarni	41 – 48
Sistem Informasi Manajemen Bantuan Untuk Korban Bencana Alam Berbasis Web Titan Parama Yoga, Iis Ismail	49 – 58
Model Sistem Pendukung Keputusan Transportasi melalui Metode Saving Matrix Pada CV XYZ Tombak Gapura Bhagya	59 – 68
Rancang Bangun Single Page Application Berbasis Framework Laravel Dan Elm (Studi Kasus E-Job XYZ) Trisna Gelar Abdullah, Budiman	69 – 78

STUDI KOMPARASI ALGORITMA SIMILARITAS PADA PREDIKSI RATING BERBASIS ITEM PADA COLLABORATIVE FILTERING STUDI KASUS PADA DATA REVIEW RESTORAN

Mochamad Iqbal Ardimansyah^{#1}

Program Studi Teknik Informatika, Universitas Islam Nusantara Bandung

¹iqbalardimansyah@gmail.com

ABSTRAK

Collaborative filtering merupakan salah satu teknik yang memanfaatkan informasi preferensi pengguna dalam bentuk penilaian peringkat (*rating*) yang menghasilkan prediksi berdasarkan kesamaan pola penilaian. Akurasi prediksi selalu menjadi penilaian pada sistem yang dibangun dengan teknik *collaborative filtering*. Studi ini melibatkan komparasi algoritma similaritas yang digunakan pada *collaborative filtering* berbasis item untuk memprediksi penilaian (*rating*) dalam studi kasus data restoran. Pengujian yang dilakukan adalah membangun sistem *collaborative filtering* berbasis item dengan menggunakan variansi algoritma similarity antara *euclidean distance* dan *cosine similarity* yang bertujuan untuk menganalisis kemampuan keduanya dalam memprediksi item. Hasil studi pada kasus ini menunjukkan bahwa *collaborative filtering* berbasis item dengan pendekatan euclidean distance memiliki akurasi yang lebih baik daripada *cosine similarity*.

Kata Kunci: studi komparasi, algoritma similaritas, collaborative filtering, prediksi, rating.

ABSTRAK

Collaborative filtering is one technique that utilizes user preference information in the form of the valuation of ratings that produce predictions based on similarity of patterns evaluation. Prediction accuracy is always an assessment of a system built with collaborative phasing techniques. This study involves the comparison of similarity algorithms used in item-based collaborative filtering to predict ratings in restaurant data case studies. The test is to build an item-based collaborative filtering system by using the variance of the similarity algorithm between euclidean distance and cosine similarity which aims to analyze the ability of both in predicting items. The results of the study in this case indicate that item-based collaborative filtering with the euclidean distance approach has better accuracy than cosine similarity.

Keywords: comparative study, similarity algorithms, collaborative filtering, prediksi, rating.

1. PENDAHULUAN

Beberapa tahun terakhir, perkembangan teknologi yang pesat telah menghasilkan ukuran data yang sangat besar, situasi ini membuat sulit dalam hal pencarian informasi yang relevan dan berguna. *Collaborative filtering* adalah salah satu teknik yang digunakan untuk sistem rekomendasi untuk memberikan prediksi item kepada pengguna dengan mengamati preferensi dari pengguna lain. Pendekatan penyaringan kolaboratif lebih umum digunakan karena lebih efisien, tidak memerlukan pengetahuan sebelumnya tentang sifat-sifat item [1].

Dalam teknik *collaborative filtering* terdiri dari beberapa proses utama diantaranya pertama tahap representasi data selanjutnya mencari kesamaan preferensi lalu kemudian memprediksi penilaian (*rating*). Keakuratan dalam memprediksi salah satunya tergantung pada metrik similaritas yang digunakan, oleh karena itu untuk mencari similaritas preferensi pada studi ini menggunakan beberapa variasi metrik algoritma similaritas antara *euclidean distance* dan *cosine similarity*. Untuk tahap selanjutnya memprediksi penilaian (*rating*) kami menggunakan *weight sum*.

Fokus pada penelitian ini adalah menganalisis bagaimana perbandingan algoritma similaritas yang digunakan pada *collaborative filtering* dengan studi kasus pada data review restoran.

2. TINJAUAN TEORI

a. Studi Komparasi

Studi dalam kamus bahasa indonesia studi adalah penelitian ilmiah, sedangkan Komparasi menurut Sjachran Basah, merupakan suatu metode pengkajian atau penyelidikan dengan mengadakan perbandingan di antara dua objek kajian atau lebih.

Sehingga dapat dikatakan studi komparasi merupakan penelitian yang bersifat membandingkan yang dilakukan dengan tujuan membandingkan persamaan dan perbedaan diantara dua atau lebih objek yang akan dijadikan kajian berdasarkan pemikiran kerangka tertentu.

b. Collaborative Filtering

Collaborative filtering (CF) merupakan bagian dari teknik sistem rekomendasi dengan mempertimbangkan persamaan preferensi pengguna lain. Dalam hal ini perilaku preferensi baik secara eksplisit seperti skor penilaian ataupun secara implisit seperti data transaksi.

Cara kerja *collaborative filtering* ini dengan cara menghimpun preferensi pengguna dalam bentuk penilaian (*rating*) dalam suatu domain dengan memanfaatkan kesamaan dan perbedaan antar profil dari beberapa pengguna untuk menentukan bagaimana memprediksi suatu item.

Table 1. Contoh Rating Pengguna dan Item

	Item 1	Item 2	Item 3	Item 4	Item 5
Adam	5	-	2	3	-
Budi	2	-	-	-	-
Arif	3	-	-	5	-
Eva	-	-	3	4	-
Kania	-	4	3	-	5

Secara umum terdapat 3 jenis *collaborative filtering*, yaitu *memory based CF*, *model based CF* dan *hybrid CF*[3]. *Memory based CF*, prediksi dihasilkan dengan mencari kesamaan pengguna atau kesamaan item. *Model based CF*, prediksi yang dihasilkan dari datasets yang kemudian akan dibuat suatu pemodelan yang berisi rating dari user, dari model ini kemudian akan dihitung similarity tiap tiap elemennya. Dan yang terakhir *hybrid based CF* merupakan kombinasi dari *memory based CF* dan *model based CF*.

c. Collaborative Filtering Berbasis Item

Collaborative filtering berbasis item menentukan prediksi atas preferensi test-pengguna terhadap sebuah test-item berdasarkan informasi rating milik item lain yang similar dengan test-item, penyajian prediksinya dengan cara mengidentifikasi hubungan antara item. dimana ketetapan dalam pendekatan ini adalah item[4].

d. Algoritma Similaritas Collaborative Filtering

Algoritma similaritas *collaborative filtering* adalah algoritma yang digunakan untuk mengukur seberapa mirip diantara dua objek

data. Jika *space* jarak rendah, maka tingkat similaritasnya tinggi begitupun sebaliknya *space* jaraknya tinggi maka tingkat similaritasnya rendah.

- Euclidean Distance

Dalam hal ini, *euclidean distance* adalah jarak antara dua buah objek atau titik. *Euclidean distance* dapat digunakan untuk mengukur kemiripan sebuah objek dengan objek yang lainnya, secara matematis dapat dituliskan di dalam persamaan berikut [5]

$$\text{sim}(i, j) = \frac{1}{\sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2}}$$

x_i = Nilai titik 1, x_j = Nilai titik 2

- Cosine Similarity

Dalam hal ini, dua item dianggap sebagai dua vektor di dalam *space* pengguna dimensi m . Similaritas diantara item-item dihitung dengan perhitungan cosinus sudut antara dua vektor. Secara formal, di dalam matriks $m \times n$, similarity antara item i dan j , dinotasikan dengan $\text{sim}(i, j)$, secara matematis dapat dituliskan di dalam persamaan berikut [6]

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

dimana \cdot adalah notasi dari dot-product dari dua vektor, dimana $\vec{i} = (i_1, i_2, \dots, i_n)$ dan $\vec{j} = (j_1, j_2, \dots, j_n)$

Pendekatan *cosine similarity* memberikan rekomendasi yang lebih baik apabila nilai $\text{sim}(i, j)$ antara 0.0 sampai 1.0. Nilai 1.0 menunjukkan bahwa kedua item adalah sama [6]

- e. *Prediksi*

Metode prediksi penilaian (*rating*) yang digunakan adalah *weight sum* dimana metode ini untuk menghitung prediksi dengan menjumlahkan penilaian (*rating*)

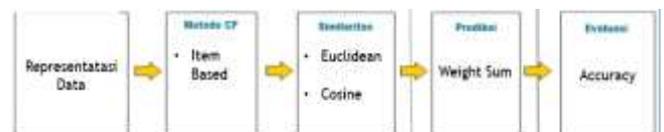
yang diberikan oleh pengguna pada item yang mirip dengan item. Saat membuat prediksi dari similaritas terdekat, tingkat similaritas dapat ditimbang berdasarkan jarak ke target, secara matematis dapat dituliskan di dalam persamaan berikut [6]

$$P_{a,i} = \frac{\sum_{u=1}^n (r_{u,i}) * w_{a,u}}{\sum_{u=1}^n w_{a,u}}$$

Di mana $r_{u,i}$ adalah peringkat pengguna tetangga u ke item target i , $w_{a,u}$ adalah kesamaan dari pengguna target dan pengguna tetangga u , dan n adalah angka tetangga.

3. DESAIN DAN IMPLEMENTASI

Desain sistem yang dibangun dengan membuat tahapan *collaborative filtering* pada umumnya hanya saja pada tahapan mencari kesamaan pada studi komprasi ini menggunakan beberapa algoritma similaritas yang berbeda. Desain sistem dapat dilihat pada Gambar. 1,



Gambar1. Desain Sistem

Pendekatan yang diusulkan mulai dari representasi data, metode *collaborative filtering*, perhitungan kesamaan (*similarity*), prediksi dan evaluasi.

Pada studi ini, dalam mencari kesamaan (*similarity*) pada teknik *collaborative filtering* kami mengusulkan komparasi antara *Euclidian Distance* dan *Cosine Similarity*.

4. EKSPERIMEN DAN HASIL

A. Skenario Eksperimen

Untuk menguji realibilitas prediksi pada studi komparasi ini, kami mengukur keakuratan hasil prediksi dengan membandingkan teknik

collaborative filtering berbasis item dengan variasi algoritma antara algoritma similaritas *euclidian distance* dengan *cosine similarity*.

Ukuran tetangga dalam perbandingan ini yang digunakan berkisar dari k-10 hingga k-30. Dalam hal ini tetangga k-N adalah pengguna yang serupa, di mana N adalah jumlah tetangga dengan nilai kesamaan tertinggi di antara pengguna lain. Misalnya, tetangga k-10 pada pengguna u adalah 10 tetangga dengan nilai kesamaan tertinggi dengan pengguna u.

B. Set Data

Data yang digunakan dari situs kaggle, data mentah adalah tabel yang terdiri dari kolom id, id item dan nilai rating (1-5). konstruksi data mentah untuk membentuk m x n dengan m adalah total pengguna dan n adalah total item, 138 pengguna pada 130 item restoran, sampel data dapat dilihat pada Gambar. 2

userID	placeID	rating	food_rating	service_rating
U1077	135085	2	2	2
U1077	135038	2	2	1
U1077	132825	2	2	2
U1077	135060	1	2	2
U1068	135104	1	1	2
U1068	132740	0	0	0
U1068	132663	1	1	1
U1068	132732	0	0	0
U1068	132630	1	1	1

Gambar2. Sampel Set Data Restoran

C. Metrik Evaluasi

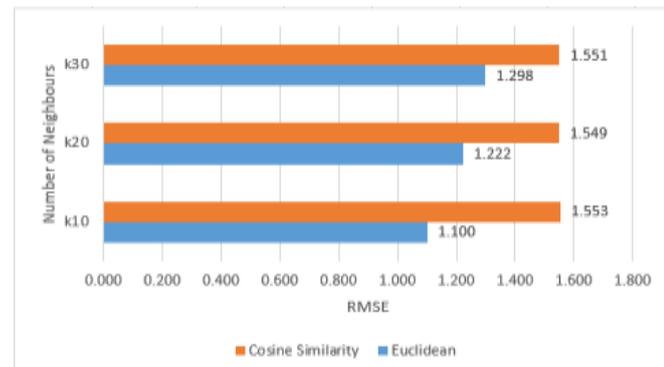
Dalam studi komparasi ini, kami menggunakan root mean square error (RMSE) untuk mengevaluasi kinerja pengaruh algoritma similaritas yang dibandingkan terhadap akurasi prediksi yang akan dihasilkan. Kumpulan peringkat prediksi hasil pengguna N adalah $\{\hat{r}_1, \hat{r}_2, \dots, \hat{r}_n\}$ akan dibandingkan dengan set penilaian sejati yang ditetapkan adalah $\{r_1, r_2, \dots, r_n\}$. Perhitungan metrik evaluasi didefinisikan sebagai berikut [8]

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (\hat{r}_j - r_j)^2}$$

Kami menggunakan RMSE sebagai metrik evaluasi untuk hasil prediksi. Semakin rendah RMSE, semakin akurat dalam memprediksi.

D. Hasil Eksperimen

Pada studi komparasi ini kami akan membandingkan *collaborative filtering* berbasis item dengan pencarian kesamaan menggunakan *euclidean distance* dengan *cosine similarity*.



Gambar3. RMSE perbandingan *collaborative filtering* berbasis item dengan algoritma similaritas *euclidean distance* dan *cosine similarity*

Berdasarkan grafik pada Gambar. 3, di semua berbagai ukuran ketetanggan 'kN' dengan pendekatan yang di komparasikan antara algoritma similaritas *euclidean distance* dan *cosine similarity* tampak bahwa algoritma dengan similaritas *euclidean distance* nilai kesalahan (RMSE) lebih kecil dari hasil dengan algoritma similaritas *cosine similarity* dengan tingkat kesalahan (error) rata-rata perbedaan 0.344, hal ini dikarenakan perbedaan cara mengukur pada masing-masing algoritma similaritas, *cosine similarity* mengukur jarak dengan mempertimbangkan sudut vektor sedangkan *euclidean distance* mengukur secara kesejajaran antara titik satu dengan yang lainnya. Perbedaan tingkat kesalahan (error) pada kedua algoritma similaritas yang komparasi ini dipengaruhi juga oleh karakteristik set data yang digunakan.

5. KESIMPULAN

Hasil prediksi dari komparasi yang dilakukan menunjukkan perbedaan tingkat kesalahan (error) prediksi pada *collaborative filtering* berbasis

item antara *euclidean distance* dan *cosine similarity*, tingkat kesalahan (error) yang rendah mengindikasikan kualitas prediksi yang baik. Penggunaan algoritma similaritas membuat dampak yang cukup besar dalam memberikan akurasi prediksi yang lebih baik.

REFERENSI

- [1] Jerome Kunegis, Andreas Lommatzsch, Martin Mehlitz, and Sahin Albayrak. "Assessing the Value of Unrated Items in Collaborative Filtering," ICDIM'07, 2nd International Conference on Digital Information Management, 2007. pp. 212-216.
- [2] Sjachran Basah, Ilmu Negara, Bandung: Citra Aditya Bhakti, 1994.
- [3] Xiaouan Su, Khoshgoofar, T.M. "A survey of collaborative filtering techniques," Advances in artificial intelligence, 2009.
- [4] Francesco Ricci and Lior Rokach and Bracha Shapira, "Introduction to Recommender Systems Handbook," Springer, 2011, pp. 1-35.
- [5] Riyadi A. . *Data Mining*. Bandung: Yrama Widya, 2009.
- [6] Claudio Adrian Levinas. "An Analysis of Memory Based Collaborative Filtering Recommender Systems with Improvement Proposals," Master of Science Thesis, 2014.
- [7] <https://www.kaggle.com/>, Restaurant Data with Consumer Ratings, 19 Januari 2019.
- [8] Schroder G, Thiele M, and Lehner. W. "Setting Goals and Choosing Metrics for Recommender System Evaluations," In: Proceedings of the Workshop on Human Decision Making in Recommender Systems and User-Centric Evaluation of Recommender Systems and Their Interfaces, CEUR Workshop Proceedings, 2011. vol. 811, pp. 78-85.