COLLABORATIVE FILTERING BERBASIS ALGORITMA PEARSON CORRELATION DAN WEIGHT AVERAGE SEBAGAI SISTEM REKOMENDASI E-COMMERCE WISATA PULAU LOMBOK

Retno Ekayanti1, Penulis Dua2, Penulis Tiga3

1,2 Jurusan Sistem Informasi Fakultas Teknologi Informasi Universitas Nasional   
Jln. Raya Cilegon Serang – Drangong Kota Serang

1retnoekayanti@unas.ac.id

2fauziah@unas.ac.id

3 Institusi yang berbeda dipisah

3email\_penulis3@instansi.ac.id

Abstrak

Sistem rekomendasi adalah sistem yang bertujuan untuk memberi saran kepada pengguna terhadap suatu item berdasarkan preferensi riwayat pengguna lain. Salah satu metode yang ada pada sistem rekomendasi adalah metode item based collaborative filtering yang beroperasi dengan mengambil informasi pengguna lain berupa nilai rating untuk direkomendasikan.

Pada penelitian ini, dibuat sebuah sistem yang dapat merekomendasikan item berupa paket wisata Lombok kepada pengguna. Sistem dibangun menggunakan item based collaborative filtering dengan persamaan pearson correlation based similarity untuk menghitung nilai kemiripan item, weighted average of deviation untuk menghitung nilai prediksi user terhadap item, dan mean absolute error (MAE) untuk menghitung nilai kesalahan prediksi. Nilai MAE yang telah dihasilkan akan diurutkan berdasarkan nilai terkecil yang kemudian direkomendasikan kepada user. Dataset yang digunakan dalam penelitian ini berasal dari data Horizon berupa 198 data users, 9 data paket tour, dan total 206 data rating.

Berdasarkan nilai MAE yang dihasilkan, didapati sistem rekomendasi yang dibuat menghasilkan rekomendasi yang cukup akurat. Hasil dari penelitian ini menunjukkan kecilnya rata-rata MAE sebesar 0,525777778 dengan lama waktu running sistem 13.457173109055 detik.

**Kata kunci***:* *Item Based Collaborative Filtering, Paket Wisata, Sistem Rekomendasi*

1. **Pendahuluan**

Pemerintah telah menetapkan Kawasan Ekonomi Khusus (KEK) Mandalika, Kabupaten Lombok Tengah, Provinsi Nusa Tenggara Barat melalui Peraturan Pemerintah Nomor 52 Tahun 2014. Dan pada Jum’at, 12 November 2021, Presiden Joko Widodo meresmikan Sirkuit Mandalika yang terletak di Kawasan Ekonomi Khusus (KEK) Mandalika tersebut(Amir et al., 2020). Pembangunan sirkuit bertaraf internasional di kawasan KEK Mandalika sangat mempengaruhi perekonomian masyarakat lokal di berbagai macam sektor terlebih pada sektor pariwisata. Hal ini membuat masyarakat lokal turut berperan dalam bidang penunjang bisnis di Kawasan KEK Mandalika sebagai pelaku bisnis pariwista(Sutanto & Kunci, 2022).

Tidak seperti Hotel dan Transportasi dengan sistem pengelolaan informasi yang baik, beberapa pengusaha paket wisata masih kurang optimal dalam menyampaikan informasi mengenai jasa yang mereka tawarkan. Hal ini berpengaruh pada ketertarikan turis asing dan lokal yang ingin berwisata menggunakan paket tour yang disediakan. Turis yang belum mengetaui lokasi wisata Lombok kesulitan dalam memilih paket yang sesuai dengan paket wisata yang mereka sukai. Cara yang selama ini dilakukan adalah turis akan berkonsultasi langsung kepada admin atau pihak wisata(Isnaini, 2020).

Namun disamping kendala terhadap pihak eksternal, pihak internal Wisata juga memiliki keterbatasan dalam mengelola data wisatawan seperti yang dialami oleh Horizon Tour and Travel. Banyaknya pesan masuk membuat pihak travel kesulitan membalas pesan satu-persatu sehingga mengakibatkan adanya kesalahan dalam membagikan jenis paket yang di tawarkan hingga proses transaksi.

Penelitian terdahulu telah memberikan solusi dari permasalahan terkait, salah satu contohnya adalah permasalahan mengenai user yang kesulitan dalam menentukan wisata kuliner yang kemudian dibuatlah sebuah sistem rekomendasi wisata kuliner. Adanya sistem ini dapat membantu user dalam mengambil keputusan mengenai apa yang diinginkan(Devi Nurhayati & Widayani, 2021). Namun belum ada sistem informasi dan media transaksi untuk user dapat menlanjutkan aktivitas mengenai apa yang telah direkomendasikan.

Berdasarkan masalah yang ada, calon wisatawan memerlukan sistem informasi dengan memanfaatkan perkembangan teknologi untuk membantu menentukan paket wisata yang akan dipilih berdasarkan hasil rekomendasi yang melibatkan nilai rating terdahulu. Sistem yang dibuat juga akan memudahkan dalam melakukan pemesanan paket wisata baiak transaksi online maupun cash on delivery. Selain itu, biro wisata juga memerlukan sistem yang dapat mengelola informasi dengan baik untuk pencatatan data wisata dan pelaporan transaksi paket wisata. Pentingnya sistem infromasi bukan hanya dalam mempermudah pengolahan data namun juga untuk meningkatkan pelayanan kepada pelanggan di tengah persaingan pihak wisata lainnya.

Collaborative filtering merupakan salah satu metode untuk menghasilkan sebuah rekomendasi kepada user berdasarkan kesamaan item yang terlah dipilih user terdahulu. Nilai rekomendasi diambil dari nilai rating pada setiap produk yang telah diberi rating. Collaborative filtering memanfaatkan 3 algoritma yang saling berelasi yaitu Pearson Correlation untuk menghitung nilai similarity, Weighted Average Of Deviation untuk menghitung nilai prediksi, dan Mean Absolute Error untuk uji akurasi.

Contoh pemanfaatan collaborative filtering secara sederhana adalah user A telah memberi rating pada 3 item yang pernah di pilih (item 1, item 2, dan item 3). User B juga telah memberi rating pada 3 item yang pernah di pilih (item 1, item 3, dan item 4). Kemudian user C ingin mengetahui item mana yang akan ia pilih setelah ia memilih item 1 dan item 4. Dengan menggunakan metode collaborative filtering maka sistem akan merekomendasikan item 3 untuk user C. Hasil tersebut didapat karena terdapat kesamaan item yang dipilih sebelumnya antara user B dan user C. Contoh lainnya adalah ketika ada user baru yang belum pernah memilih item apapun ingin mengetahui rekomendasi apa yang ditawarkan, maka digunakanlah perhitungan akhir menggunakan mean absolute error untuk menguji akurasi prediksi tiap tiap user yang telah memberi rating.

1. **Metodologi Penelitian**

*Collaborative Filtering* adalah salah satu metode yang dapat digunakan pada bidang rekomendasi e-commerce denganmemanfaatkan rekomendasi yang diberikan berdasarkan pertimbangan data dari pengguna lain(Wayan Priscila Yuni Praditya et al., 2021). Pengimplementasiannya tidak mengharuskan pengguna untuk secara aktif memberikan informasi tentang kebutuhan pribadi mereka, tetapi memperoleh preferensi potensial mereka berdasarkan catatan *rating*  yang ada(Lin et al., 2022).

1. Algoritma Pearson Correlation Based Similarity

Pearson Correlation Based Similarity merupakan persamaan yang digunakan untuk menghitung kemiripan item dengan mempertimbangkan adanya nilai kosong yang tidak dimiliki user. Koefisien korelasi Pearson digunakan untuk menguji korelasi linier antara dua variabel. Koefisien diukur dalam skala tanpa satuan dan dapat mengambil nilai dari -1 hingga 0 hingga +1. Nilai yang mendekati nol menyebutkan tidak ada korelasi linier dan nilai yang mendekati +1 atau -1 menyiratkan korelasi linier yang sempurna(Faroqi et al., 2020).

Berikut adalah persamaan algoritma adjusted cosine similarity:

(1)

merupakan *similarity* antara *tour* *package* dan *tour* *package* , dan adalah rating rata-rata pada paket tour dan , dan adalah rating oleh user kepada paket tourdan , lalu merupakan jumlah total *user* yang memberi rating.

1. Weighted Average Of Deviation

Weight Average of Deviation merupakan perhitungan prediksi score dari berbagai item dengan membandingkan kemiripan rating yang diberikan oleh user(Prasetyo et al., 2019).

Dibawah ini adalah rumus dari persamaan algotima weight sum:

(2)

Persamaan (2) menunjukkan dimana P\_(u,k) adalah prediksi rating paket tour k untuk user u, n adalah jumlah user, R ̅\_k adalah rating rata-rata pada paket tour k, R ̅\_l adalah rating rata-rata pada paket tour 1, R\_(u,l) adalah rating yang diberikan user u kepada paket tour l, dan Sim(k,l) adalah nilai similarity antara paket tour k dengan seluruh user yang memberi rating pada paket tour ke-l.

1. Mean Absolute Error (MAE)

Mean absolute error (MAE) adalah algoritma untuk menguji akurasi sistem rekomendasi. Semakin kecil nilai MAE maka semakin akurat prediksi rating dari sebuah sistem rekomendasi(Devi Nurhayati & Widayani, 2021). Persamaan MAE dapat dilihat pada persamaan (3).

(3)

Persamaan (3) menunjukkan bahwa P\_(u,k) merupakan prediksi rating terhadap user u untuk item k, R\_(u,k) adalah nilai rating yang diberikan oleh user u untuk item k, kemudian N adalah jumlah user yang ada.

1. **Hasil dan Pembahasan**

Hasil penelitian hendaknya dituliskan secara jelas dan padat. Diskusi hendaknya menguraikan arti pentingnya hasil penelitian, bukan mengulanginya. Hindari penggunaan sitasi dan diskusi yang berlebihan tentang literatur yang telah dipublikasikan.

1. Format Penulisan

Ukuran kertas harus sesuai dengan ukuran halaman A4, yaitu 210 mm (8,27") lebar dan 297 mm (11,69") panjang. Batas margin ditetapkan sebagai berikut:

* Atas = 19 mm (0,75")
* Bawah = 43 mm (1,69")
* Kiri = Kanan = 14,32 mm (0,56")

Artikel penulisan harus dalam format dua kolom dengan ruang 4,22 mm (0,17") antara kolom.

Paragraf harus teratur. Semua paragraf harus rata, yaitu sama-sama rata kiri dan dan rata kanan.

1. Jumlah Halaman

Jumlah halaman paper yang diajukan ke Jurnal Sistem Informasi (JSiI) adalah **antara 5 sampai dengan 7 halaman.**

1. Huruf-huruf Dokumen

Seluruh dokumen harus dalam Times New Roman atau Times font dan konten artikel menggunakan font size 10 pt. Type 3 font tidak boleh digunakan. Jenis font lain dapat digunakan jika diperlukan untuk tujuan khusus.

1. Judul dan Penulis

Judul harus dalam tipe Reguler 24 pt font. Nama pengarang harus dalam Reguler 11 pt font. Afiliasi penulis harus dalam Italic 10 pt. Alamat email harus di 10 pt Regular font.

Judul dan pengarang harus dalam format kolom tunggal dan harus terpusat. Setiap awal kata dalam judul harus huruf besar kecuali untuk kata-kata pendek seperti, "sebuah", "dan", "di", "oleh", "untuk", "dari", , "pada", "atau", dan sejenisnya.  
Penulisan penulis tidak boleh menunjukkan nama jabatan (misalnya Dosen Pembimbing), apapun gelar akademik (misalnya Dr) atau keanggotaan dari setiap organisasi profesional (misalnya Senior Member IEEE).

Agar tidak membingungkan, nama keluarga ditulis di bagian terakhir dari masing-masing nama pengarang (misalnya Yohanes AK Smith). Setiap afiliasi harus termasuk, setidaknya, nama perusahaan dan nama negara tempat penulis didasarkan (misalnya Kausal Productions Pty Ltd, Australia).

Alamat email wajib bagi penulis korespondensi (*corresponding author*). Keterangan sebagai penulis korespondensi dituliskan setelah alamat email.

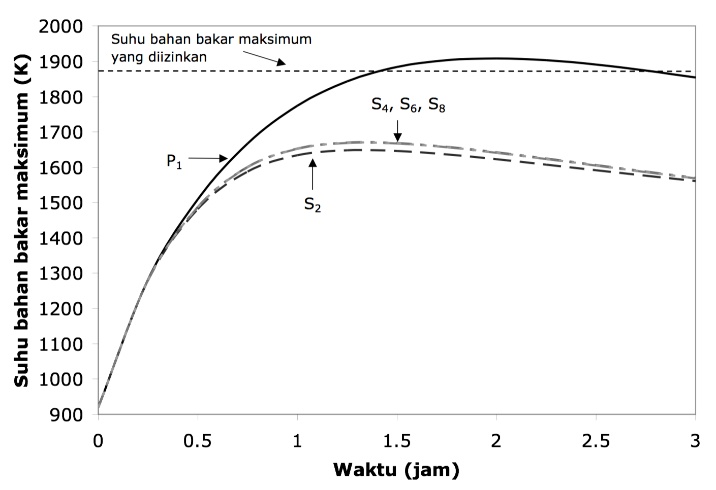
1. Bagian Heading

Sebaiknya tidak lebih dari 3 tingkat untuk *heading*. Semua tulisan harus dalam font 10pt. Setiap awal kata dalam judul *heading* harus huruf besar kecuali untuk kata-kata pendek seperti yang tercantum dalam Bagian III-D.

1. Heading Level-1: Heading level 1 harus dalam Small Caps, terletak di tengah-tengah dan menggunakan penomoran angka Romawi huruf besar. Sebagai contoh, lihat judul "III. Hasil dan Pembahasan" dari dokumen ini. Judul dengan heading 1 yang tidak diberi nomor adalah "Ucapan Terima Kasih" dan "Referensi".
2. Heading Level-2: Heading level-2 harus ditulis miring, merapat ke kiri dan nomor menggunakan abjad huruf besar. Sebagai contoh, lihat judul "E. Bagian Heading" di atas.
3. Heading Level-3: Heading level-3 harus ditulis menjorok, miring dan dinomori dengan angka Arab diikuti dengan tanda kurung kanan. Heading level 3 harus diakhiri dengan titik dua. Tulisan bersambung mengikuti judul heading dengan baris yang sama. Sebagai contoh, ayat ini diawali dengan heading level 3.
4. Gambar dan Keterangan Gambar

Gambar harus terletak di tengah (*centered*). Gambar yang besar bisa direntangkan di kedua kolom. Setiap gambar yang mencakup lebih dari 1 kolom lebar harus diposisikan baik di bagian atas atau di bagian bawah halaman. Gambar tidak diberi bingkai (*border)* di luar bidang gambar.

Gambar grafik dimungkinkan berwarna. Untuk grafik berwarna, pastikan warna cukup kontras untuk membedakan garis yang satu dengan yang lain. Untuk grafik hitam putih, gunakan jenis garis yang berbeda (misalnya garis utuh, garis putus-putus, garis titik-titik, dan sebagainya).



**Gambar 1**. Contoh Grafik Garis Menggunakan Hitam-Putih. Perhatikan penulisan label dan satuan pada sumbu horisontal maupun vertikal.

Label pada sumbu horisontal maupun vertikel sering kali dapat membingungkan. Oleh karena itu sedapat mungkin gunakan kata-kata, dan bukan hanya simbol. Berikan satuan di dalam tanda kurung, namun penulisan label jangan hanya berupa satuan tanpa keterangan secukupnya. Contoh penulisan label yang tepat adalah ”Suhu (K)” atau ”Suhu, Tmax (K)”. Contoh penulisan label yang keliru adalah ”Tmax” atau ”(K)”.

Gunakan font yang konsisten dan seragam pada grafik. Font yang disarankan adalah Times New Roman (atau Times), Arial (atau Helvetica), Symbol dan Courier. Contoh grafik dapat dilihat pada Gambar 1.

Pastikan bahwa resolusi gambar cukup untuk mengungkapkan rincian penting pada gambar tersebut. Untuk gambar yang bersumber dari file JPG, pastikan mempunyai resolusi sebesar 300 dpi. Gambar 2 menunjukkan contoh sebuah gambar dengan resolusi rendah yang kurang sesuai ketentuan, sedangkan Gambar 3 menunjukkan contoh dari sebuah gambar dengan resolusi yang memadai.

Keterangan gambar diletakkan di bagian bawah gambar. Keterangan gambar menggunakan 8 pt Reguler font dan diberi nomor dengan menggunakan angka Arab. Keterangan gambar dalam satu baris (misalnya Gambar 2) diletakkan di tengah (*centered*), sedangkan keterangan gambar yang lebih dari satu baris harus dirata kiri (misalnya Gambar 1). Keterangan gambar dengan nomor gambar harus ditempatkan sesuai dengan poin-poin yang relevan, seperti ditunjukkan pada Gambar 1 – 3, kecuali jika gambar berukuran besar melebihi satu kolom.



**Gambar 2**. Contoh gambar dengan resolusi kurang

1. Tabel dan Keterangan Tabel

Tabel harus diletakkan di tengah (*centered*). Tabel yang besar bisa direntangkan di kedua kolom atau diputar menjadi vertikal. Setiap tabel atau gambar yang mencakup lebih dari 1 kolom lebar harus diposisikan baik di bagian atas atau di bagian bawah halaman.

Tabel dan judul tabel ditulis dengan font 8 pt Reguler. Tabel diberi nomor menggunakan angka Romawi huruf besar. Setiap awal kata dalam judul tabel menggunakan huruf besar kecuali untuk kata-kata pendek seperti yang tercantum pada Bagian III-D, dan ditulis menggunakan Small Caps. Keterangan angka tabel ditempatkan sebelum tabel terkait. Isi tabel ditulis rata tengah. Perhatikan bahwa pada tabel tidak terdapat garis vertikal. Contoh tabel dapat dilihat di Tabel I.

**Tabel 1**. Potensi Konversi Beberapa Radionuklida

|  |  |  |  |
| --- | --- | --- | --- |
| **Nuklida** | **Energi Neutron** | | |
| **Termal** | **Epitermal** | **Cepat** |
| Pu-239 | 1,09 | 0,9 | 1,9 |
| U-233 | 1,20 | 1,3 | 1,5 |
| U-235 | 1,07 | 0,8 | 1,3 |

1. Persamaan

Persamaan ditulis rata tengah. Gunakan *Microsoft Equation Editor* atau *MathType add-on*. Jangan *copy paste* persamaan dari file lain yang berbentuk pdf atau jpg. Penomoran persamaan ditulis rata kanan dengan angka arab di dalam tanda kurung. Contoh penulisan persamaan dapat dilihat di Persamaan (1) berikut ini.

  (1)

Gunakan ukuran (*size*) berikut ini pada *Microsoft Equation Editor*:

* Full : 10 pt
* Subscript/Superscript : 5 pt
* Sub-subscript/superscript : 3 pt
* Symbol : 16 pt
* Sub-symbol : 10 pt

1. Nomor Halaman, Headers dan Footers

Nomor halaman, headers dan footers tidak dipakai.

1. Tautan dan Bookmark

Semua tautan *hypertext* dan bagian *bookmark* akan dihapus. Jika paper perlu merujuk ke alamat email atau URL di artikel, maka alamat atau URL lengkap harus ditulis dengan font biasa.

1. Referensi

Setiap dokumen/pustaka yang disitasi di paper ini harus dituliskan di bagian ini. Jumlah pustaka yang disitasi minimal 10 buah, dengan 80% berupa acuan primer. Yang dimaksud dengan acuan primer adalah artikel jurnal, book chapter, paten, paper seminar/prosiding. Adapun yang dimaksud dengan acuan sekunder adalah buku teks dan handbook.

Judul pada bagian Referensi tidak boleh bernomor. Semua item referensi dalam 8 pt font. Silakan gunakan Italic Reguler dan gaya untuk membedakan berbagai bidang seperti ditunjukkan pada bagian Referensi. Nomor referensi harus ditulis berurutan di dalam tanda kurung siku (misalnya [1]).  
Ketika mengacu pada item referensi, silakan menggunakan nomor referensi saja, seperti dalam [2]. Jangan menggunakan "Ref. [3]" atau "Referensi [3]" kecuali pada awal kalimat, misalnya "Referensi [3] menunjukkan bahwa ...". Pengacuan beberapa referensi dengan cara menuliskan masing-masing nomor dengan kurung terpisah (misalnya [2], [3], [4] - [6]). Beberapa contoh item referensi dengan kategori yang berbeda ditampilkan dalam bagian Referensi meliputi:

* contoh untuk buku pada [1]
* contoh sebuah buku dalam seri dalam [2]
* contoh artikel jurnal di [3]
* contoh paper seminar di [4]
* contoh paten dalam [5]
* contoh website di [6]
* contoh dari suatu halaman web di [7]
* contoh manual databook dalam [8]
* contoh datasheet di [9]
* contoh tesis master di [10]

1. **Kesimpulan**

Kesimpulan berisi tentang poin-poin utama artikel. Kesimpulan hendaknya tidak mengulangi yang sudah dituliskan di bagian Intisari, akan tetapi membahas tingkat keberhasilan dari penerapan maupun pengembangan dari penelitian yang dilakukan. Bagian ini hendaknya juga dapat menunjukkan apakah tujuan penelitian dapat tercapai.

Kesimpulan ditulis dalam bentuk paragraf uraian. Hindari penggunaan *bulleted list*.

**V. Saran**

Bagian ini memberikan saran kepada penulis lain yang membaca atau menjadikan artikel ini sebagai salah satu referensi. Saran dapat dijadikan ide/gagasan/pengembangan penelitian selanjutnya.

**Referensi**

1. S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.
2. J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.
3. S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, “A novel ultrathin elevated channel low-temperature poly-Si TFT,” *IEEE Electron Device Lett.*, vol. 20, pp. 569–571, Nov. 1999.
4. M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, “High resolution fiber distributed measurements with coherent OFDR,” in *Proc. ECOC’00*, 2000, paper 11.3.4, p. 109.
5. R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, “High-speed digital-to-RF converter,” U.S. Patent 5 668 842, Sept. 16, 1997.
6. (2002) The IEEE website. [Online]. Available: http://www.ieee.org/
7. M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/
8. *FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.
9. “PDCA12-70 data sheet,” Opto Speed SA, Mezzovico, Switzerland.
10. A. Karnik, “Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP,” M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.
11. J. Padhye, V. Firoiu, and D. Towsley, “A stochastic model of TCP Reno congestion avoidance and control,” Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.
12. *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.