



E-ISSN 3032-601X & P-ISSN 3032-7105

Vol. 1, No. 3c, Juli 2024

MISTER

Journal of Multidisciplinary Inquiry in Science,
Technology and Educational Research

**Jurnal Penelitian Multidisiplin dalam Ilmu
Pengetahuan, Teknologi dan Pendidikan**

**UNIVERSITAS SERAMBI MEKKAH
KOTA BANDA ACEH**

mister@serambimekkah.ac.id

Journal of Multidisciplinary Inquiry in Science Technology
and Educational Research

Journal of MISTER

Vol. 1, No. 3c, Juli 2024

Pages: 1560–1566

Sistem Rekomendasi Linimasa Facebook Berdasarkan Topik Kesukaan Pengguna Menggunakan Metode Content-Based Filtering & Term Frequency-Inverse Document Frequency

Bayu Setiawan, Bagus Satrio Wicaksono, Muhammad Hilmy Aziz, Anggraini Puspita Sari

Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional
“Veteran” Jawa Timur, Surabaya, Indonesia

Article in Journal of MISTER

Available at : <https://jurnal-serambimekkah.org/index.php/mister/index>

DOI : <https://doi.org/10.32672/mister.v1i3c.2004>

How to Cite this Article

APA : Setiawan, B., Wicaksono, B. S., Aziz, M. H., & Sari, A. P. (2024). Sistem Rekomendasi Linimasa Facebook Berdasarkan Topik Kesukaan Pengguna Menggunakan Metode Content-Based Filtering & Term Frequency-Inverse Document Frequency . *MISTER: Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research*, 1(3c), 1560 – 1566.
<https://doi.org/10.32672/mister.v1i3c.2004>

Others Visit : <https://jurnal-serambimekkah.org/index.php/mister/index>

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is a scholarly journal dedicated to the exploration and dissemination of innovative ideas, trends and research on the various topics include, but not limited to functional areas of Science, Technology, Education, Humanities, Economy, Art, Health and Medicine, Environment and Sustainability or Law and Ethics.

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is an open-access journal, and users are permitted to read, download, copy, search, or link to the full text of articles or use them for other lawful purposes. Articles on Journal of MISTER have been previewed and authenticated by the Authors before sending for publication. The Journal, Chief Editor, and the editorial board are not entitled or liable to either justify or responsible for inaccurate and misleading data if any. It is the sole responsibility of the Author concerned.



ISSN 3032-7105

ISSN 3032-601X



9

773032

710001

9

773032

601002

Sistem Rekomendasi Linimasa Facebook Berdasarkan Topik Kesukaan Pengguna Menggunakan Metode Content-Based Filtering & Term Frequency-Inverse Document Frequency

Bayu Setiawan¹, Bagus Satrio Wicaksono², Muhammad Hilmy Aziz³,
Angraini Puspita Sari⁴

Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur, Surabaya, Indonesia^{1,2,3,4}

Email Korespondensi: anggraini.puspita.if@upnjatim.ac.id

Diterima: 11-07-2024

| Disetujui: 13-07-2024

| Diterbitkan: 14-07-2024

ABSTRACT

This research aims to identify Facebook timeline recommendation systems based on user habits using Content-Based Filtering (CBF) and Term Frequency-Inverse Document Frequency (TF-IDF) methods. By utilizing historical user activity data, this system will customize content recommendations for individual users. The CBF method is used to compare the similarity of the content to be recommended with the user's habits, while TF-IDF is used to evaluate various keywords in the content. The results of this study show that the combination of CBF and TF-IDF methods can improve the accuracy of recommendation in Facebook timeline, resulting in a more convenient and relevant user experience.

Keywords: Recommendation Systems, Facebook Timeline, Content-Based Filtering, Term Frequency, Inverse Document Frequency.

ABSTRAK

Penelitian ini bertujuan untuk mengidentifikasi sistem rekomendasi linimasa Facebook berdasarkan kebiasaan pengguna menggunakan metode *Content-Based Filtering* (CBF) dan *Term Frequency-Inverse Document Frequency* (TF-IDF). Dengan memanfaatkan data historis aktivitas pengguna, sistem ini akan menyesuaikan rekomendasi konten untuk pengguna secara individual. Metode CBF digunakan untuk membandingkan kemiripan konten yang akan direkomendasikan dengan kebiasaan pengguna, sedangkan TF-IDF digunakan untuk mengevaluasi berbagai kata kunci di dalam konten. Hasil dari penelitian ini menunjukkan bahwa kombinasi metode CBF dan TF-IDF dapat meningkatkan akurasi rekomendasi linimasa Facebook, menghasilkan pengalaman pengguna yang lebih nyaman dan relevan.

Katakunci: Sistem Rekomendasi, Linimasa Facebook, Content-Based Filtering, Term Frequency, Inverse Document Frequency.

PENDAHULUAN

Platform media sosial seperti Facebook telah menjadi bagian yang cukup penting dalam kehidupan sehari-hari di masyarakat, sebagai sumber informasi, hiburan, dan komunikasi (Manasijević, Živković, Arsić, & Milošević, 2016). Dengan jumlah konten pasti bertambah setiap harinya, pengguna sering mendapati konten yang kurang diminati yang dianggap tidak relevan dengan preferensi mereka, hal ini menjadi salah satu faktor mengapa pengguna berpaling dari media sosial (Kosinski, Matz, Gosling, Popov, & Stillwell, 2015). Namun, tidak mudah untuk untuk memproses data dalam jumlah besar secara efektif, sehingga diperlukan sistem seperti *machine learning* yang dapat belajar dari data ini. Untuk mengatasi masalah ini, sistem rekomendasi konten harus dibuat sebagai alat untuk meningkatkan pengalaman pengguna dengan memberikan rekomendasi konten yang relevan dan menarik sesuai dengan preferensi individual pengguna (Javed et al., 2021). Salah satu faktor kunci dalam pengembangan sistem rekomendasi konten yang efektif adalah dapat memahami perilaku dan preferensi pengguna. Dengan menganalisis interaksi pengguna, seperti suka, komentar, berbagi, dan riwayat penelusuran, dari kebiasaan tersebut dapat diperoleh preferensi, minat dan kebiasaan pengguna dengan cukup akurat.

Dengan memanfaatkan informasi pengguna yang telah didapat, sistem rekomendasi dapat secara efektif untuk menyaring dan memprioritaskan konten yang lebih menarik dan relevan untuk disajikan ke pengguna. Dalam konteks ini, penelitian ini berfokus pada identifikasi rekomendasi konten untuk linimasa Facebook berdasarkan kebiasaan pengguna menggunakan kombinasi metode *Content-Based Filtering* (CBF) dan *Term Frequency-Inverse Document Frequency* (TF-IDF). CBF memanfaatkan kesamaan antara item konten dan preferensi pengguna, sementara TF-IDF mengevaluasi setiap kata kunci dalam konten dan setiap kata kunci akan diberikan bobot tersendiri, bobot dari setiap kata kunci inilah yang digunakan untuk mengurutkan konten yang paling relevan dan kurang relevan kepada pengguna, semakin besar bobot yang dimiliki maka akan semakin besar juga tingkat kemiripan konten dengan preferensi pengguna.

Kelemahan dari metode *Content-based filtering* adalah kecenderungan hanya menyarankan *item* yang sangat mirip dengan *item* yang telah disukai atau dinilai tinggi oleh pengguna sebelumnya (Mondi, Wijayanto, & Winarno, 2019). Akibatnya, rekomendasi yang dihasilkan mungkin kurang bervariasi dan bahkan tidak memperkenalkan pengguna kepada konten baru yang berbeda namun masih berpotensi menarik bagi mereka.

Pada sebuah penelitian untuk mendapatkan suatu rekomendasi produk (Shrivastava & Sisodia, 2019). Penelitian tersebut menggunakan nama produk dengan bahasa Inggris sebagai kueri suatu produk dan menguji pada produk lainnya yang kemudian dicari kemiripan antara keduanya, dengan menggunakan Teknik *bag of word*, nama produk akan diubah ke matriks yang menyimpan kata unik beserta jumlahnya. Bisa dikatakan, TF-IDF menyediakan kata kunci dengan beberapa dokumen tertentu sehingga dapat diidentifikasi atau dikategorikan (Shahzad & Ramsha, 2018). Kata unik yang sudah diubah ke matriks akan diberi bobot menggunakan metode TF-IDF

Dengan mengintegrasikan metode tersebut, sistem dapat meningkatkan akurasi dan relevansi konten, sehingga dapat meningkatkan pengalaman pengguna secara keseluruhan di platform Facebook. Artikel ini memaparkan metodologi, implementasi, dan evaluasi dari pendekatan yang diusulkan, menyoroti efektivitasnya dalam mengidentifikasi rekomendasi konten personal untuk pengguna Facebook. Hasil dari penelitian ini bertujuan pada upaya yang sedang berlangsung untuk mengembangkan sistem rekomendasi

yang lebih canggih dan berorientasi pengguna untuk platform media sosial, dengan tujuan akhir meningkatkan kepuasan dan keterlibatan pengguna.

METODE PENELITIAN

a) Pengumpulan Data

Pengumpulan data merupakan proses atau tahapan yang bertujuan untuk mendapatkan informasi berdasarkan sumber yang valid, misal: jurnal ilmiah, repositori, website, buku, dataset, dan sebagainya. Hal tersebut berguna sebagai pendukung dalam suatu penelitian (Sihananto et al., 2022).

Data yang digunakan dalam studi kasus ini berasal dari *data world* (data world, 2016), yang merupakan sebuah platform *dataset* berbasis komunitas serta *open source*. *Dataset* tersebut terdiri dari total 20 kolom, sementara yang digunakan pada studi kasus ini hanya 5 kolom, yaitu judul posting, isi posting, total *likes*, total komen, dan tanggal posting. Data inilah yang nantinya akan diolah.

b) Data Preprocessing

Sebelum melanjutkan ke tahap perhitungan TF-IDF, dilakukan penyaringan data untuk menangani nilai yang hilang. Data yang hilang bisa terjadi karena berbagai alasan seperti penghapusan konten atau kesalahan saat pengumpulan data, dengan menggunakan metode sederhana mengganti *missing values* dengan *string* kosong. Mengisi nilai yang hilang dengan *string* kosong mencegah kesalahan saat mencoba memproses teks menjadi vektor TF-IDF. Dengan mengisi nilai hilang, dapat dipastikan setiap entri memiliki nilai *string* yang dapat diproses.

c) Term Frequency – Inverse Document Frequency (TF-IDF)

Term Frequency (TF) mewakili frekuensi kata, yang menunjukkan berapa kali kata tersebut muncul dalam *dataset* (persamaan 1). Perhitungan yang dilakukan yaitu dengan menghitung kemunculan kata dari jumlah total kata yang ada di dalam *dataset*.

$$tf_i = \frac{n_i}{\sum_k n_k} \text{ (persamaan 1)} \quad idf_i = \log \frac{|D|}{|\{d_j: t_i \in d_j\}|} \text{ (persamaan 2)}$$

Inverse Document Frequency (IDF) mengukur seberapa penting kata dalam *dataset* secara keseluruhan. Proses perhitungannya yaitu dengan menghitung logaritma kebalikan dari proporsi kata dalam *dataset* yang memuat kata tersebut (persamaan 2). Hasil logaritma tadi merupakan nilai dari IDF. Bobot dari TF-IDF dihitung dengan mengalikan kedua ukuran tersebut ($tf_i * idf_i$), sehingga semakin besar bobotnya, semakin penting kata yang tersebut (Septiani & Isabela, 2022).

d) Content-Based Filtering

Content-based filtering menyediakan rekomendasi personal kepada pengguna berdasarkan karakteristik *item* dan preferensi pengguna sebelumnya. Hal tersebut dilakukan dengan melakukan penghitungan *similarity* atribut yang terdapat dalam item (Liu, Xu, & Tang, 2019). Fungsi *similarity* yang paling banyak digunakan adalah *adjusted cosine similarity*, *cosine similarity*, dan *pearson*

coefficient. Pada sistem studi kasus ini, akan digunakan fungsi *cosine similarity* yang dapat mengukur kemiripan antara konten yang satu dengan konten yang lainnya (Saraswathi et al., 2022). *Cosine similarity* akan menganggap suatu kata sebagai vektor.

Algoritma *Cosine Similarity* mengukur kesamaan dengan menggunakan sudut kosinus antara kedua vektor A dan B, yang dapat mewakili kata-kata, kalimat, paragraf atau bahkan seluruh dokumen (persamaan 3). Semakin dekat nilai cosinusnya dengan 1 maka semakin kecil sudut antara dua vektor. Jika menerapkan algoritma ini ke *textual corpora*, maka dapat digunakan untuk mengevaluasi tingkat kemiripan antar dokumen (Pramesti & Santiyasa, 2022).

$$\cos \theta = \frac{A.B}{||A|| ||B||} \text{ (persamaan 3)}$$

HASIL DAN PEMBAHASAN

Hal pertama yang dilakukan untuk mendapatkan rekomendasi linimasa facebook adalah menghitung TF-IDF, metode ini menghitung seberapa besar bobot yang di dapat pada suatu postingan. semakin besar bobot yang di dapat dari proses tersebut maka postingan akan muncul paling awal. Bobot ditentukan dengan kesesuaian postingan dengan kebiasaan pengguna. Setelah itu akan dimasukkan ke dalam array *cosine similarity* dan data yang di dapat dari TF-IDF akan dihitung, Dari perhitungan TF-IDF dan *cosine similarity* berdasarkan kebiasaan pengguna dengan keyword sebagai berikut.

No	Keyword
1	boxing
2	technology
3	space

dari keyword yang telah ada didapatkan data sebagai berikut.

Title	Message	Comments Count	Likes	Date	Score
Face Off: Mayweather/Cotto	In a sit down with Max Kellerman and Floyd Mayweather, Miguel Cotto explains that his fight with Mayweather will be the biggest and most important one of his carrer. Mayweather and Cotto address questions about Manny Pacquiao and discuss their upcoming fight ahead of tonight's HBO Boxing series 24/7 MAYWEATHER/COTTO,	69	306	22-04-2016	0.374

	which airs on HBO at 10 ET/PT and re-airs on CNN at midnight.				
Portrait of a fighter: Floyd Mayweather	“You know, when I’m out there, it’s ‘Money Mayweather,’ ‘Human ATM Machine,’ I mean it’s different names they call me. But you know, when I come home, I’m either ‘Dad’ or just ‘Floyd.’” In advance of tomorrow night’s HBO Boxing 24/7 MAYWEATHER/COTTO installment airing on HBO at 10 ET/PT and re-airing on CNN at midnight, Floyd Mayweather opens up about his turbulent childhood, family and what it means to be a champion. How does this change your view of the fighter?	160	345	21-04-2016	0.161
CNN video on Facebook	One day this weekend, the moon will appear larger and brighter than any other full moon in 2016. http://on.cnn.com/IKz4xJ Find out why and which day it will happen in your area.	137	674	05-05-2016	0.077
So you're a cyborg -- now what?	If you've ever found yourself running late because you left your phone at home, 'you might be a cyborg.'	101	379	08-05-2016	0.0775
Quirky.com CEO on The Next List	This week The Next List will be profiling Ben Kaufman, founder and CEO of Quirky - check out the preview and don't miss the show today at 2pm ET	31	157	06-05-2016	0.775

Berdasarkan data tabel postingan berikut, diketahui bahwa skor yang didapat akan lebih besar jika kata kunci lebih relevan pada suatu postingan. Hal ini akan menentukan postingan mana yang akan ditampilkan ke pengguna.

KESIMPULAN

Kesimpulan dari penelitian ini adalah salah satu cara untuk menentukan postingan atau konten yang relevan kepada pengguna adalah dengan metode TF-IDF dengan *content based filtering*, dan dari perhitungan yang sudah dijabarkan terbukti bahwa metode TF-IDF bisa digunakan untuk menentukan postingan atau konten yang relevan kepada pengguna dengan menghitung bobot dari setiap keyword yang ada. Setelah bobot ditemukan, akan diurutkan mulai dari bobot terbesar hingga terkecil. Dari hasil tersebut menunjukkan bahwa aplikasi ini sudah terbilang efektif, hanya saja belum dapat memberi hasil bobot yang lebih tinggi dikarenakan keterbatasan pada *dataset* yang ada.

DAFTAR PUSTAKA

- Manasijević, D., Živković, D., Arsić, S., & Milošević, I. (2016). Exploring students' purposes of usage and educational usage of Facebook. *Computers in Human Behavior*, 60, 441-450.
- Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, 70(6), 543-556.
- Shrivastava, R., & Sisodia, D. S. (2019). Product recommendations using textual similarity based learning models. In *2019 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-7). Coimbatore, India.
- Shahzad, Q., & Ramsha, A. (2018). Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *International Journal of Computer Applications ER*, 181(1), 25.
- Javed, U., Shaukat, K., Hameed, I. A., Iqbal, F., Alam, T. M., & Luo, S. (2021). A review of content-based and context-based recommendation systems. *International Journal of Emerging Technologies in Learning (iJET)*, 16(3), 274-306.
- Steck, H., Baltrunas, L., Elahi, E., Liang, D., Raimond, Y., & Basilico, J. (2021). Deep learning for recommender systems: A Netflix case study. *AI Magazine*, 42(3), 7-18.
- Gomez-Urbe, C. A., & Hunt, N. (2015). The Netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, 6(4), 1-19.
- Mondi, R. H., Wijayanto, A., & Winarno, W. (2019). Recommendation system with content-based filtering method for culinary tourism in Mangan application. *ITSMART: Jurnal Teknologi dan Informasi*, 8(2), 65-72.
- Chiny, M., Chihab, M., Bencharef, O., & Chihab, Y. (2022). Netflix recommendation system based on TF-IDF and cosine similarity algorithms. In *International Conference on Big Data, Machine Learning, and Applications (BML)* (pp. 15-20). Rabat, Morocco.
- Sihananto, A. N., Sari, A. P., Khariono, H., Fernanda, R. A., & Wijaya, D. C. M. (2022). Implementasi metode K-Means untuk pengelompokan kasus Covid-19 tingkat provinsi di Indonesia. *Jurnal Informatika dan Sistem Informasi*, 3(1), 76-85.
- Septiani, D., & Isabela, I. (2022). Analisis term frequency inverse document frequency (tf-idf) dalam temu kembali informasi pada dokumen teks. *Sistem dan Teknologi Informasi Indonesia (SINTESIA)*, 1(2), 81-88.
- Liu, Y., Xu, Q., & Tang, Z. (2019). Research on text classification method based on PTF-IDF and cosine similarity. In *2019 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)* (pp. 205-208). Shanghai, China.

- Saraswathi, K., Mohanraj, V., Suresh, Y., & Senthilkumar, J. (2022). Deep learning enabled social media recommendation based on user comments. *Computer Systems Science & Engineering*, 44(2), 1691.
- Pramesti, D. A. P. D., & Santiyasa, I. W. (2022). Penerapan metode content-based filtering dalam sistem rekomendasi video game. *Jurnal Nasional Teknologi Informasi dan Aplikasinya*, 1(1), 229–234.
- Turcotte, J., York, C., Irving, J., Scholl, R. M., & Pingree, R. J. (2015). News recommendations from social media opinion leaders: Effects on media trust and information seeking. *Journal of Computer-Mediated Communication*, 20(5), 520–535.