
Implementasi TF-IDF, *Cosine Similarity*, dan *Logistic Regression* Pada Rekomendasi Buku Berdasarkan *Mood* Pembaca Dengan Data Oversampling

Intan Sanu^{1)*}, Jeovanni Wong²⁾ & Hafiz Irsyad³⁾

^{1,2,3)}Prodi Informatika, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang, Indonesia

*Corresponding Email: intansanu_2226250109@mhs.mdp.ac.id

Abstrak

Pembaca seringkali kesulitan dalam memilih buku yang sesuai dengan *mood* atau suasana hati mereka. Sistem rekomendasi buku yang dapat menyesuaikan dengan *mood* diharapkan dapat meningkatkan kepuasan pembaca. Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi buku berdasarkan *mood* pembaca menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)*, *Cosine Similarity*, dan *Logistic Regression*. Penelitian ini menggunakan dataset buku yang diambil dari situs *Kaggle* yang berjumlah 6.810 data dan dibersihkan melalui tahap *pre-processing* untuk menghilangkan data yang tidak lengkap. Sistem ini menggunakan TF-IDF untuk mengubah deskripsi buku menjadi representasi numerik, serta *Cosine Similarity* untuk menghitung kesamaan antar buku berdasarkan deskripsi dan *mood* pengguna. Hasil penelitian menunjukkan bahwa sistem dapat memberikan rekomendasi buku yang relevan, dengan tingkat akurasi yang mencapai 92%, dengan nilai *precision* sebesar 93%, *recall* sebesar 89%, dan *F1-score* sebesar 91%. Evaluasi dilakukan dengan menggunakan *confusion matrix*, yang menunjukkan hasil yang cukup memuaskan dengan *precision*, *recall*, dan *F1-score* yang baik. Penelitian ini menunjukan potensi dalam penggunaan teknologi dalam meningkatkan pengalaman membaca dengan menyediakan rekomendasi yang sesuai dengan kondisi emosional pembaca.

Kata Kunci: *Cosine Similarity*, *Mood* pembaca, Rekomendasi Buku, TF-IDF, *Logistic Regression*.

Abstract

Readers often struggle to find books that match their mood or emotional state. A mood-based book recommendation system is expected to enhance reader satisfaction. This study aims to develop a book recommendation system based on the reader's mood using *Term Frequency-Inverse Document Frequency (TF-IDF)*, *Cosine Similarity*, and *Logistic Regression*. The study utilizes a dataset of 6.810 books sourced from *Kaggle*, which was cleaned through *pre-processing* to remove incomplete data. The system uses TF-IDF to convert book descriptions into numerical representations and applies *Cosine Similarity* to measure the similarity between books based on their descriptions and the user's mood. The results show that the system is capable of providing relevant book recommendations, achieving an accuracy of 92%, with a precision of 93%, recall of 89%, and an *F1-score* of 91%. Evaluation using a *confusion matrix* indicates satisfactory performance, with strong precision, recall, and *F1-score* values. This research demonstrates the potential of leveraging technology to enhance the reading experience by offering recommendations aligned with the reader's emotional state.

Keywords: *Book Recommendation*, *Cosine Similarity*, TF-IDF, *reader's mood*, *Logistic Regression*.

PENDAHULUAN

Pada era digital saat ini, pembaca menghadapi berbagai pilihan buku yang sangat luas, mencakup berbagai genre, tema, hingga gaya penulisan. Namun demikian, pemilihan buku yang sesuai dengan suasana hati atau *mood* pembaca seringkali menjadi tantangan tersendiri. Suasana hati memainkan peran penting dalam menentukan preferensi seseorang terhadap materi tertentu. Oleh karena itu, memiliki sistem rekomendasi yang dapat menyesuaikan dengan suasana hati atau *mood* pembaca berpotensi untuk meningkatkan kepuasan serta pengalaman membaca secara keseluruhan.

Sistem rekomendasi umumnya mengandalkan pendekatan berbasis konten (*content-based*) atau *collaborative filtering* yang belum mempertimbangkan aspek emosional pengguna. Sedangkan, integrasi aspek psikologis seperti *mood* ke dalam sistem rekomendasi dapat menghasilkan hasil yang lebih personal. Salah satu pendekatan yang dapat digunakan untuk membangun sistem rekomendasi semacam itu adalah kombinasi metode *Term Frequency-Inverse Document Frequency (TF-IDF)*, *Cosine Similarity*, dan *Logistic Regression*. TF-IDF berperan dalam menentukan kata kunci dalam dokumen, sementara *Cosine Similarity* digunakan untuk menghitung tingkat kesamaan antar dokumen berdasarkan vektor representasinya.

Beberapa studi menunjukkan efektivitas kombinasi TF-IDF dan *Cosine Similarity* dalam sistem rekomendasi. Permana dan Wibowo (2023) mengembangkan sistem rekomendasi film berdasarkan sinopsis menggunakan kedua metode tersebut untuk mengukur relevansi konten. Hal yang serupa juga dilakukan oleh Rymmai dan Saleema (2017), yang memanfaatkan ringkasan cerita untuk menghitung kesamaan antar buku. Sementara itu, Ardiansyah et al. (2023) menerapkan metode *content-based filtering* untuk merekomendasikan buku perpustakaan berdasarkan minat siswa menggunakan TF-IDF dan *Cosine Similarity*.

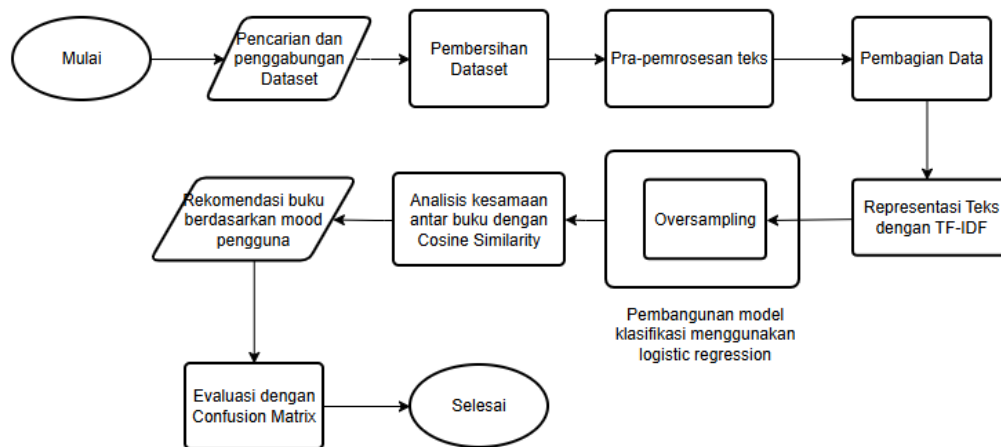
Di sisi lain, Rahmadhani et al. (2024) menerapkan pembobotan TF-RF dan *Cosine Similarity* untuk meningkatkan akurasi pencarian buku digital.

Meskipun pendekatan tersebut terbukti efektif dalam konteks kesamaan konten, aspek emosional pembaca, seperti suasana hati, masih jarang diterapkan dalam sistem rekomendasi. Padahal, memiliki potensi besar dalam menghadirkan pengalaman membaca yang lebih relevan dan memuaskan. Oleh karena itu, penelitian ini bertujuan mengembangkan sistem rekomendasi buku berbasis *mood* pembaca dengan memanfaatkan kombinasi TF-IDF, *Cosine Similarity*, dan *Logistic Regression*. Dengan menerapkan metode ini, sistem dapat memberikan rekomendasi buku yang sesuai dengan suasana hati pembaca, sehingga dapat meningkatkan pengalaman dan kepuasan pembaca.

METODE PENELITIAN

Metode yang digunakan dalam penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk membangun dan melakukan pengujian sistem rekomendasi buku berdasarkan *mood* pembaca dengan menggunakan teknik pemrosesan teks TF-IDF dan *cosine similarity*. Menurut Creswell (2014) pendekatan kuantitatif adalah pendekatan untuk menguji teori-teori tertentu dengan cara menguji hubungan antar variabel. Variabel-variabel ini dapat diukur, biasanya dengan instrumen, sehingga data angka dapat dianalisis menggunakan prosedur statistik. Sugiyono (2016) penelitian eksperimen diartikan sebagai metode penelitian yang digunakan untuk mencari pengaruh perlakuan tertentu terhadap yang lain dalam kondisi yang terkendalikan.

Metode ini terbagi dari beberapa tahapan, dimulai dari pencarian dan penggabungan dataset, pembersihan dataset, pra-pemrosesan teks, pembagian data, representasi teks dengan TF-IDF, pembangunan model klasifikasi menggunakan *logistic regression*, analisis kesamaan antara buku dengan *cosine similarity*, sistem rekomendasi buku, dan evaluasi.



Gambar 1. Tahapan Metode Penelitian

Dimulai dengan mencari dataset buku dari situs *Kaggle*, dataset buku yang didapat berisi sebanyak 6.810 data dengan kolom *isbn13*, *isbn10*, *title*, *subtitle*, *authors*, *categories*, *thumbnail*, *description*, *published_year*, *average_rating*, *num_pages*, dan *ratings_count*. Adapun dataset tambahan yang berisi *isbn13*, *title*, dan *true_mood* yang diperoleh dari file *test-ground-truth.csv* yang berisi sebanyak 200 data buku dari dataset *kaggle*. Nilai *true_mood* diperoleh dari pencarian secara manual berdasarkan judul buku, kemudian digabungkan dengan dataset utama berdasarkan kolom *title*. Kolom tersebut digunakan sebagai *ground truth* dalam pelatihan dan evaluasi model klasifikasi *mood*.

Setelah penggabungan, dataset dibersihkan dengan menghapus data yang memiliki nilai kosong pada kolom *description* dan *true_mood*. Dataset yang memiliki kurang dari dua data per kategori *mood* akan dihapus guna untuk menjaga keseimbangan kelas saat pelatihan model.

Tahap pra-pemrosesan teks bertujuan untuk membersihkan dan menstandarkan data teks yang akan digunakan dalam pemodelan. Pada tahap ini data *description* buku digunakan untuk pemrosesan teks menggunakan *library*

TfidfVectorizer yang akan secara otomatis melakukan pemrosesan teks. Adapun tahapannya dimulai dari mengubah teks menjadi *lowercase*, menghapus *stopwords*, tanda baca dan simbol non-kata, lalu tokenisasi yaitu proses memecah teks menjadi kata.

Setelah tahap pra-pemrosesan dilakukan, dataset dibagi menjadi 80% data latih dan 20% data uji. Pembagian dataset bertujuan agar model dapat mempelajari pola dari data latih. Data latih digunakan untuk membangun model klasifikasi *mood*, sedangkan data uji untuk mengevaluasi seberapa baik model dalam menggeneralisasi data baru.

Tahapan representasi teks dengan TF-IDF, dilakukan dengan mengubah deskripsi buku menjadi representasi numerik. Representasi numerik dilakukan agar sistem dapat menghitung kesamaan antar buku berdasarkan isi deskripsi secara matematis agar dapat merekomendasikan buku lain dengan *mood* serupa. TF digunakan untuk mengukur seberapa sering kata tertentu muncul dalam satu dokumen dibandingkan dengan jumlah seluruh kata dalam dokumen, TF dihitung dengan rumus.

$$TF(t, d) = \frac{\text{Jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{Jumlah kata dalam dokumen } d} \quad (1)$$

Kemudian dilakukan perhitungan IDF untuk mengukur seberapa penting kata tersebut di dokumen.

$$IDF(t) = \log \frac{\text{Total jumlah dokumen dalam } d}{\text{Jumlah dokumen dimana kata } t \text{ muncul}} \quad (2)$$

Pembangunan model klasifikasi menggunakan *logistic regression* sebagai model klasifikasi untuk memprediksi *mood* berdasarkan deskripsi buku. Sebelum pelatihan, dilakukannya *oversampling* pada data latih untuk mengatasi ketidakseimbangan kelas pada data *true_mood*. Kemudian model dilatih menggunakan data latih yang sudah diproses dan distandarisasi dengan TF-IDF dan *oversampling*. Setelahnya, diterapkannya *pipeline* untuk menggabungkan tahapan

TF-IDF, *oversampling*, dan *logistic regression*. Proses latih dilakukan dengan *pipeline.fit()* menggunakan data latih *x_train* dan *y_train*.

Pada tahap analisis kesamaan antar buku, input *mood* pengguna akan dihitung TF-IDF nya dengan model yang sama untuk membandingkan secara konsisten dengan TF-IDF deskripsi buku. Setelahnya, dilakukan perhitungan *cosine similarity* antara vektor TF-IDF input pengguna dengan vektor TF-IDF setiap deskripsi buku untuk menentukan kesamaannya.

$$\text{Cosine Similarity}(x, y) = \frac{x \cdot y}{||x|| ||y||} \quad (3)$$

Keterangan :

$x \cdot y$ = hasil perkalian titik antara dua vektor x dan y .

$||x|| ||y||$ = norma dari vektor x dan y .

Pada tahap rekomendasi buku berdasarkan *mood* pengguna, sistem mengambil skor dengan nilai *cosine similarity* tertinggi antara input mood pengguna dengan deskripsi buku. Buku dengan nilai cosine similarity tertinggi akan dianggap paling relevan sesuai dengan *mood* dan direkomendasikan kepada pengguna. Sehingga pengguna akan mendapatkan hasil rekomendasi buku sesuai *mood*.

Evaluasi kinerja model klasifikasi *mood* dilakukan dengan menggunakan *confusion matrix*, merupakan metode evaluasi untuk mengukur kinerja sistem rekomendasi. *Confusion matrix* terdiri dari *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil deteksi sistem (*mood*) dibandingkan dengan hasil deteksi dari kategori dan deskripsi buku yang sudah ada (*true_mood*). Jika hasil *mood* sesuai dengan *true_mood* maka prediksi dianggap benar, sedangkan jika kebalikannya maka prediksi akan dianggap salah.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$F1 - Score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \quad (7)$$

Keterangan :

True Positive (TP) = data positif yang diprediksi benar.

True Negative (TN) = data negatif yang diprediksi benar.

False Positive (FP) = data negatif namun diprediksi sebagai data positif.

False Negative (FN) = data positif namun diprediksi sebagai data negatif.

Accuracy digunakan untuk mengukur seberapa akuratnya prediksi sistem itu benar secara keseluruhan. *Precision* untuk mengukur seberapa akuratnya prediksi yang benar sesuai dengan kategori. *Recall* digunakan untuk mengukur seberapa baik sistem dapat menangkap setiap *mood* yang ada. *F1-Score* untuk menyeimbangkan antara *precision* dan *recall*, *Confusion matrix* dapat dilihat pada tabel 1.

Tabel 1. Confusion Matrix

		Prediksi	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

HASIL DAN PEMBAHASAN

Data diambil dari dataset pada *Kaggle.com* yang berjudul “7k Books” yang bertujuan untuk sistem rekomendasi buku berdasarkan *mood* pembaca. Pada dataset ini terdapat sebanyak 6.810 judul buku, dan data yang digunakan akan di *pre-processing* terlebih dahulu. Pada tahap *pre-processing*, 1.000 data dipilih secara acak untuk digunakan sebagai model, yang kemudian digabungkan dengan dataset “*ground truth*” yang memuat label *true_mood*. Setelah data bersih, dataset dibagi menjadi 80% data pelatihan dan 20% data pengujian. Dalam tahap pelatihan, metode TF-IDF digunakan untuk mengubah deskripsi teks menjadi representasi numerik, yang kemudian dilanjutkan dengan penerapan *random oversampling* untuk menangani ketidakseimbangan kelas, serta *logistic regression* digunakan sebagai model untuk klasifikasi.

Setelah model dilatih, dilakukan prediksi terhadap deskripsi buku di dataset gabungan. Kinerja model kemudian dievaluasi menggunakan *classification report* dan *confusion matrix*. *Classification report* memberikan metrik evaluasi seperti *precision*, *recall*, dan *f1-score*, yang menunjukkan kualitas prediksi model pada setiap kategori *mood*. Sedangkan, *confusion matrix* menunjukkan gambaran visual yang lebih jelas tentang seberapa akurat model memprediksi *mood* berdasarkan deskripsi buku.

Selain itu, *cosine similarity* digunakan untuk memberikan rekomendasi buku berdasarkan input *mood*. Metode ini digunakan untuk mengukur seberapa mirip input *mood* dengan deskripsi buku yang ada dalam dataset.

Masukkan mood yang Anda rasakan saat ini (contoh: romantic, dark, inspiring, thrilling):
Mood Anda: dark

Rekomendasi buku dengan mood 'dark' (urut berdasarkan kemiripan deskripsi):

Judul: Quivers
True Mood: dark | Predicted: dark
Skor Kemiripan: 0.1095
Deskripsi: The outspoken co-host of The Howard Stern Show reveals her disturbing background as a sexually abuse...

Judul: Fragile Things
True Mood: dark | Predicted: dark
Skor Kemiripan: 0.0831
Deskripsi: A mysterious circus terrifies an audience for one extraordinary performance before disappearing into...

Judul: Well of Darkness
True Mood: dark | Predicted: dark
Skor Kemiripan: 0.0581
Deskripsi: Gareth is just a frightened young lad when he is drafted in to be the whipping-boy for the unruly Pr...

Gambar 1. Rekomendasi Buku berdasarkan Mood Dark

Masukkan mood yang Anda rasakan saat ini (contoh: romantic, dark, inspiring, thrilling):
Mood Anda: thrilling

Rekomendasi buku dengan mood 'thrilling' (urut berdasarkan kemiripan deskripsi):

Judul: Appointment with Death
True Mood: thrilling | Predicted: thrilling
Skor Kemiripan: 0.0688
Deskripsi: A repugnant Amercian widow is killed during a trip to Petra... Among the towering red cliffs of Petr...

Judul: The Listerdale Mystery
True Mood: thrilling | Predicted: thrilling
Skor Kemiripan: 0.0675
Deskripsi: A selection of mysteries, some light-hearted, some romantic, some very deadly... Twelve tantalizing ...

Judul: Mrs McGinty's Dead
True Mood: thrilling | Predicted: thrilling
Skor Kemiripan: 0.0552
Deskripsi: An old widow is brutally killed in the parlour of her cottage... Mrs McGinty died from a brutal blow...

Gambar 2. Rekomendasi Buku berdasarkan Mood Thrilling

Hasil evaluasi dengan menggunakan metrik-metrik performa seperti, *precision*, *recall*, dan *f1-score* untuk setiap kelas yang diuji. Dalam Evaluasi ini, model mengklasifikasi beberapa kategori, yaitu "dark", "happy", "inspirational", "philosophical", "romantic", dan "thrilling". Kelas "happy" dan "romantic" menunjukkan hasil yang baik dengan *precision* sebesar 1.00 (100%) dan *recall* yang tinggi sebesar 0.97 (97%), menghasilkan *f1-score* yang sangat baik. Kelas "philosophical" dan "thrilling" memiliki kinerja yang baik dengan nilai *precision* sebesar 0.90 (90%) dan *recall* yang cukup tinggi sebesar 0.90 (90%), meskipun tidak sebaik kelas "happy." Kelas "inspirational" dan "dark" memiliki sedikit penurunan dalam nilai *precision* sebesar 0.89 (89%) dan *recall* sebesar 0.86 (86%), namun tetap menunjukkan kinerja yang baik. Secara keseluruhan, model ini

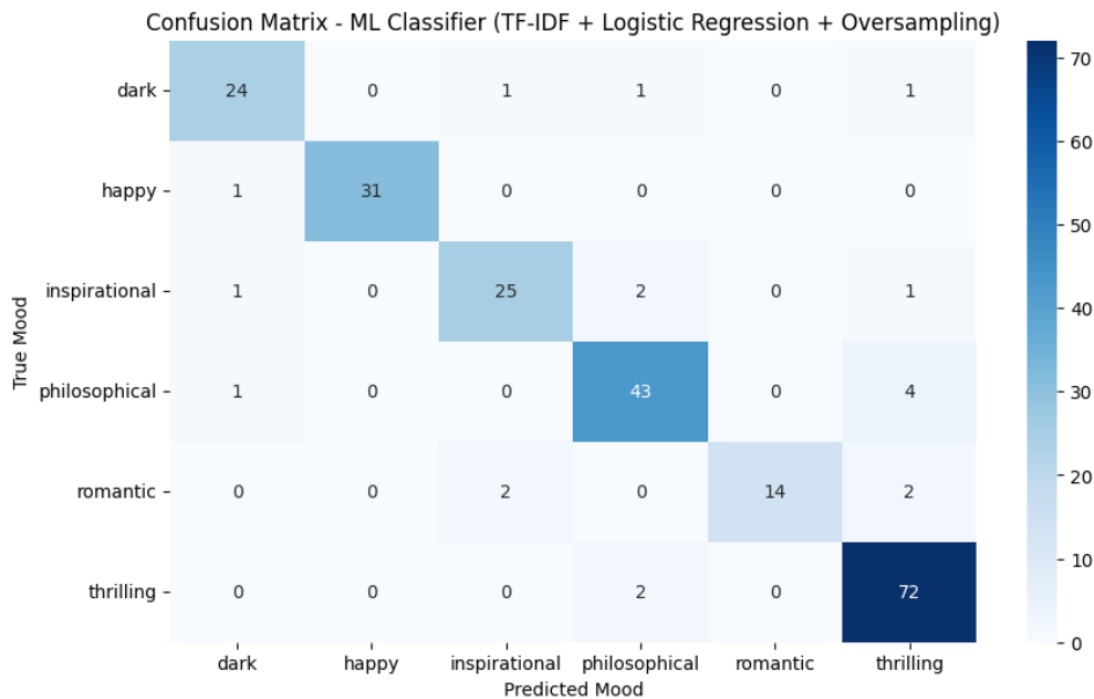
memiliki akurasi sebesar 0.92, menunjukkan bahwa 92% dari prediksi yang dilakukan model adalah benar. Rata-rata per kategori menunjukkan *precision* 0.93, *recall* 0.89, dan *f1-score* 0.91, sedangkan rata-rata berbobot mencatatkan nilai *precision*, *recall*, dan *f1-score* masing-masing 0.92. Evaluasi ini mencerminkan kinerja model yang *solid* dengan hasil yang konsisten di seluruh kategori yang diuji. Hasil dari evaluasi dapat dilihat Pada gambar 3.

=== Evaluasi Machine Learning Classifier ===

	precision	recall	f1-score	support
dark	0.89	0.89	0.89	27
happy	1.00	0.97	0.98	32
inspirational	0.89	0.86	0.88	29
philosophical	0.90	0.90	0.90	48
romantic	1.00	0.78	0.88	18
thrilling	0.90	0.97	0.94	74
accuracy			0.92	228
macro avg	0.93	0.89	0.91	228
weighted avg	0.92	0.92	0.92	228

Gambar 3. *Classification Report*

Confusion matrix untuk model klasifikasi menggunakan TF-IDF, *cosine similarity*, dan *logistic regression*. Matriks ini menggambarkan hasil prediksi model terhadap 6 kategori *mood*, yaitu “happy”, “dark”, “inspirational”, “philosophical”, “romantic”, dan “thrilling”. Matriks ini menunjukkan performa model dalam mengklasifikasi *mood*. Setiap baris dalam matriks menunjukkan kelas sebenarnya (*True Mood*), sedangkan setiap kolom menampilkan hasil prediksi dari model (*Predicted Mood*). Angka tinggi pada diagonal utama, seperti 31 pada “happy” dan 72 pada “thrilling”, menandakan banyaknya prediksi yang tepat. Meskipun ada beberapa kesalahan, model secara umum cukup akurat dalam mengenali *mood* yang benar. Hasil keseluruhan dapat dilihat Pada gambar 4.

Gambar 4. *Confusion Matrix*

SIMPULAN

Penelitian ini menunjukkan bahwa sistem rekomendasi buku berdasarkan *mood* pengguna berhasil dikembangkan dengan TF-IDF, *cosine similarity*, dan *logistic regression*. Sistem mampu mengubah deskripsi buku menjadi representasi numerik dengan TF-IDF dan berhasil mengklasifikasi *mood* menggunakan *logistic regression*. *Cosine similarity* berhasil dalam memberikan rekomendasi buku dari dataset yang sesuai dengan input *mood* pengguna. Sistem terbukti mampu menampilkan hasil buku yang relevan dengan input *mood* pengguna. Keberhasilan sistem diukur dari hasil evaluasi dengan *confusion matrix* yang menunjukkan tingkat akurasi yang tinggi dengan nilai *precision*, *recall*, dan *F1-Score* yang seimbang, menunjukkan performa sistem yang baik dalam menemukan buku yang sesuai dengan *mood* pengguna. Sehingga sistem ini sangat relevan bagi pengguna yang ingin membaca buku berdasarkan *mood*. Namun, penelitian ini masih dapat dikembangkan lebih lanjut dengan memperluas dataset dan memperbanyak label *true_mood* agar penyesuaian *mood* terhadap deskripsi buku menjadi lebih akurat.

DAFTAR PUSTAKA

- Ardiansyah, R., Ari Bianto, M. & Saputra, B.D. (2023). *Sistem Rekomendasi Buku Perpustakaan Sekolah menggunakan Metode Content-Based Filtering*. Jurnal Computer Science and Information Technology, 4(2): 510-518. Diunduh di <https://ejurnal.umri.ac.id/index.php/coscitech/article/view/5131/2462>
- Budianto, A.G., Rusilawati., Suryo, A.T.E., Cahyono, G.R., Zulkarnain, A.F. & Martunus. (2024). *Perbandingan Performa Algoritma Support Vector Machine(SVM) dan Logistic Regression untuk Analisis Sentimen Pengguna Aplikasi Retail di Android*. Jurnal Sains dan Informatika, 10(2): 1-10. Diunduh dari <https://jsi.politala.ac.id/index.php/JSI/article/view/911/202>
- Chiny, M., Chibab, M., Bencharef, O., & Chibab, Y. (2022). *Netflix Recommendation System based on TF-IDF and Cosine Similarity Algorithms*. Proceedings of the International Conference on Information Systems and Artificial Intelligence, 3101. Diunduh di https://www.researchgate.net/publication/360856267_Netflix_Recommendation_System_based_on_TF-IDF_and_Cosine_Similarity_Algorithms
- Jumadi, M., Mardian, M. & Yanti, L. (2020). *Pengaruh strategi pembelajaran The Power of Two untuk meningkatkan kemampuan membaca teks eksposisi*. Journal of Educational Review and Research, 3(1): 59-64. Diunduh di https://www.google.com/url?sa=i&url=https%3A%2F%2Fjournal.stkipsingkawang.ac.id%2Findex.php%2FJERR%2Farticle%2Fdownload%2F2064%2F1368&psig=AOvVaw38HRb5LH_vlpl7272aM5Km&ust=1747327203517000&source=images&cd=vfe&opi=89978449&ved=0CAYQrpoMahcKEwig5aubs6ONAxUAAAAAHQAAAAAQBA
- Permana, A.H.J. & Wibowo, A.T. (2023). *Movie Recommendation System Based on Synopsis Using Content-Based Filtering with TF-IDF and Cosine Similarity*. International Journal on Information and Communication Technology, 9(2): 1-14. Diunduh di <https://socjs.telkomuniversity.ac.id/ojs/index.php/ijoict/article/view/747/391>
- Rahmadhani, S., Hakim, L. & Wibowo, G. H. (2024). *Sistem Rekomendasi Penelusuran Buku Berbasis Content-Based Filtering dengan Pembobotan TF-RF*. Jurnal Informatika Polinema, 10(4): 491-500. Diunduh di <https://jurnal.polinema.ac.id/index.php/jip/article/view/5565/3939>
- Ridhwanullah, D., Kumarahadi, Y.K. & Raharja, B.D. (2024). *Content-Based Filtering pada Sistem Rekomendasi Buku Informatika*. Jurnal Ilmiah Sinus, 22(2): 57-66. Diunduh di https://www.p3m.sinus.ac.id/jurnal/index.php/e-jurnal_SINUS/article/view/840/pdf
- Rymmai, R.G. & JS, S. (2017). *Book Recommendation using Cosine Similarity*. International Journal of Advanced Research in Computer Science, 8(3): 276-281. Diunduh di <https://www.ijarcs.info/index.php/Ijarcs/article/view/2995/2978>
- Siroj, R.A., Afgani, W., Fatimah, F., Septaria, D., Zahira, G., & Salsabila, S. (2024). *Metode penelitian kuantitatif : Pendekatan ilmiah untuk analisis data*. Jurnal Review Pendidikan dan Pengajaran, 7(3), 11279-11289. Diunduh di https://www.google.com/url?sa=i&url=https%3A%2F%2Fjournal.universitaspahlawan.ac.id%2Findex.php%2Fjrpp%2Farticle%2Fdownload%2F32467%2F21663%2F106592&psig=AOvVaw0n0aY_fYp7uYDjD5NX_AdA&ust=1747326257004000&source=images&cd=vfe&opi=89978449&ved=0CAYQrpoMahcKEwig1oXYr6ONAxUAAAAAHQAAAAAQBA

Wardana, N.S., Aditiawan, F.P. & Sari, A.P. (2024). *Logistic Regression Classification with TF-IDF and FastText for Sentiment Analysis of LinkedIn Reviews*. Journal of Visions and Ideas, 4(3): 1359-1370.

Diunduh

di

https://www.researchgate.net/publication/383378392_Logistic_Regression_Classification_with_TF-IDF_and_FastText_for_Sentiment_Analysis_of_LinkedIn_Reviews

Widianto, A., Pebriyanto, E., Fitriyanti & Marna. (2024). *Document Similarity using Term Frequency-Inverse Document Frequency Representation and Cosine Similarity*. Journal of Dinda, 4(2): 149-153.

Diunduh

di

https://www.researchgate.net/publication/383453541_Document_Similarity_Using_Term_Frequency-Inverse_Document_Frequency_Representation_and_Cosine_Similarity

