

Rekomendasi Pemilihan Sepeda Motor Menggunakan Metode Content Based Filtering Dan Item Based Colaborative Filtering

Raihan Martin Permana¹, Asep Id Hadiana², Puspita Nurul Sabrina³

^{1,2,3}Informatika, Sains dan Informatika, Universitas Jenderal Achmad Yani, Cimahi Indonesia,
Kode Pos 40513

raihan.martin@students.unjani.ac.id^{1*}, asep.hadiana@lecture.unjani.ac.id²,
puspita.sabrina@lecture.unjani.ac.id³

Abstrak

Sepeda motor menjadi andalan moda transportasi yang paling dibutuhkan orang tua dan remaja saat ini. Pemilihan sepeda motor sebagai alat transportasi yang paling diminati saat ini didasari oleh keunggulan sepeda motor itu sendiri dalam hal perawatan, bahan bakar, waktu tempuh yang lebih cepat, dan kemampuan mengatasi kemacetan yang terjadi di jalan kota. peneliti melakukan penggabungan dari kedua metode yang pertama yaitu content-based filtering dimana metode ini berdasarkan preferensi pengguna berdasarkan interaksi dengan data atau informasi selanjutnya akan dicocokkan dengan serangkaian karakteristik atribut dari barang yang akan direkomendasikan. Barang dengan tingkat kecocokan tertinggi akan menjadi rekomendasi untuk pengguna untuk metode yang kedua yaitu item based collaborative filtering metode ini menghitung kemiripan item yang disuaki oleh pengguna dengan menggunakan cosine similarity kemudian item yang memiliki nilai kesamaannya disimpan dan digunakan sebagai nilai untuk menghasilkan suatu rekomendasi. evaluasi pada penelitian ini menggunakan precision, recall dan f1-score yang berguna untuk mengukur kinerja dari sistem rekomendasi. Hasil dari evaluasi memberikan hasil precision 54.5 %, recall 97.0% dan f1-score 69.6% Dengan hasil ini untuk sistem rekomendasi sepeda motor dengan menggunakan metode content based filtering dan item based collaborative filtering memiliki hasil baik berdasarkan hasil evaluasi dan dapat diimplementasikan kedalam sistem rekomendasi yang dapat membantu pengguna dalam mencari sepeda motor yang sesuai dengan kriteria dan kebutuhan pengguna.

Kata kunci: Content Based Filtering, Item Based Collaborative Filtering, Sepeda motor

A. Pendahuluan

Sepeda motor merupakan kendaraan beroda dua yang memiliki setang sebagai pengendali sepeda motor, jok sebagai penompang tubuh dan peredam guncangan, lampu sebagai alat pencahayaan, dan mesin sebagai penggerak. Sepeda motor merupakan salah satu alat transportasi yang paling penting di dunia, karena mudah digunakan dan cukup menggunakan bahan bakar yang berupa bensin sebagai bahan utama agar sepeda motor bisa berjalan. Sepeda motor menjadi andalan moda transportasi yang paling dibutuhkan orang tua dan remaja saat ini. Pemilihan sepeda motor sebagai alat transportasi yang paling diminati saat ini didasari oleh keunggulan sepeda motor itu sendiri dalam hal perawatan, bahan bakar, waktu tempuh yang lebih cepat, dan kemampuan mengatasi kemacetan yang terjadi di jalan kota. (Intakoris, 2015)

Menurut data Asosiasi Industri Sepeda Motor Indonesia (AISI), penjualan kendaraan

bermotor domestik mencapai 505.670 unit pada Mei 2024, mengindikasikan penurunan sekitar 20,64% dari bulan sebelumnya. Penjualan kendaraan bermotor juga mengalami peningkatan, naik 45% dari bulan kebulan menjadi 47.449 unit di bulan yang sama. Pada tahun 2024, pasar domestik didominasi oleh motor scooters, dengan kontribusi sebesar 90,34% dari total penjualan dan 46,24% dari total pendapatan. Sebaliknya, motor sport menyumbang 4,95% penjualan domestik dan 27,06% ekspor, Sigit Kumala, Ketua Bidang Komersial AISI, menyatakan bahwa meskipun daya beli masyarakat secara umum mengalami penurunan, namun pasar kendaraan bermotor domestik di Indonesia masih terus bertumbuh, dengan penjualan domestik mencapai kurang lebih 500 milyar unit. (Penjualan et al., 2024)

Berdasarkan penelitian terdahulu menjelaskan, banyak produsen motor yang terus berupaya meningkatkan kualitas dan menyediakan model-model baru untuk menarik pelanggan. Sebagai hasilnya, sekarang ada banyak jenis motor yang tersedia. Kecenderungan ini sering kali membuat konsumen semakin antusias dalam memilih kendaraan yang paling sesuai dengan kebutuhan mereka. Untuk mengatasi hal tersebut, konsumen membutuhkan sebuah sistem rekomendasi yang dapat membantu dalam proses pembelian kendaraan bermotor. Salah satu pendekatan yang dapat digunakan adalah *item-based collaborative filtering*, yang memberikan rekomendasi berdasarkan masukan dari konsumen mengenai berbagai jenis kendaraan bermotor. Aplikasi ini dikembangkan dengan menggunakan metode *waterfall*. Hasil penelitian menunjukkan bahwa sistem rekomendasi bekerja dengan baik dan dapat digunakan secara komprehensif. Aplikasi ini memudahkan konsumen untuk memilih jenis skuter yang tepat.

Selain itu terdapat penelitian terdahulu yang telah dilakukan menjelaskan dalam penelitian ini menggunakan metode *content-based filtering*. Perangkat lunak yang dikembangkan memiliki tujuan untuk memudahkan pengguna dalam menentukan spesifikasi produk yang sesuai untuk motor yang akan dibeli. Sistem yang digunakan pada penelitian ini adalah aplikasi berbasis web yang menggunakan PHP dan MySQL sebagai bahasa pemrogramannya. Sistem yang sudah dirancang dapat memberikan rekomendasi sepeda motor berdasarkan spesifikasi yang diinginkan pengguna.

Berdasarkan penelitian terdahulu diatas, peneliti melakukan penggabungan dari kedua metode yang pertama yaitu *content-based filtering* dimana metode ini berdasarkan preferensi pengguna berdasarkan interaksi dengan data atau informasi selanjutnya akan dicocokkan dengan serangkaian karakteristik atribut dari barang yang akan direkomendasikan. Barang dengan tingkat kecocokan tertinggi akan menjadi rekomendasi untuk pengguna (Fajriansyah et al., 2021) untuk metode yang kedua yaitu *item based collaborative filtering* metode ini menghitung kemiripan item yang disukai oleh pengguna dengan menggunakan *cosine similarity* kemudian item yang memiliki nilai kesamaannya disimpan dan digunakan sebagai nilai untuk menghasilkan suatu rekomendasi. (Jaja et al., 2020)

B. Metode

Pengumpulan Data

Data yang digunakan pada penelitian rekomendasi pemilihan sepeda motor ini berbahasa Inggris yang di dapatkan dari situs Kaggle. Penggunaan data dari situs tersebut memiliki cukup banyak informasi mengenai data sepeda motor, situs tersebut juga menyediakan sumber daya data yang luas yang dapat diunduh oleh para pengguna. Sumber daya ini mencakup berbagai dataset dari berbagai bidang dan topik yang dapat digunakan untuk analisis data.

Preprocessing Data

Pada tahapan ini merupakan fungsi dari suatu perubahan data sepeda motor yang sudah tersedia datanya untuk diproses menjadi data yang siap diolah menjadi objek penelitian dengan melakukan proses seperti

1. *Data Cleaning*
2. *Data Selection*
3. *Tokenizing*
4. *stopword removal*

Implementasi Content Based Filtering

Content-Based Filtering adalah sistem rekomendasi berdasarkan preferensi pengguna berdasarkan interaksi dengan data atau informasi. Algoritma yang digunakan dalam *content based filtering* menyarankan item dengan membandingkannya dengan item yang diindikasikan positif oleh pengguna di masa lalu. Item terbaik kemudian direkomendasikan dengan membandingkan hasil dari item tersebut dengan item lainnya. (Arfisko, 2022)

Content-based filtering dapat merekomendasikan sepeda motor dengan cara mengelola data yang digunakan sebagai query adalah hasil dari *preprocessing data* berupa gabungan dari atribut yang terdapat dari dataset yang digabungkan dalam atribut *processed_text* lalu menghitung bobot TF-IDF untuk setiap atribut yang terdapat didalam dataset.

TF-IDF adalah teknik untuk menghasilkan bobot dari setiap kata baru untuk memberikan wawasan penting tentang kata-kata yang digunakan dalam dokumen. Metode ini sering digunakan dalam text mining. TF memiliki fungsi untuk menyoroti kata-kata yang sering muncul dalam sebuah dokumen. Namun, hal ini tidak berlaku untuk keseluruhan dokumen. IDF merupakan inversi dari Document Frequency (DF) berfungsi untuk dokumen yang memiliki persamaan. (Raharjo et al., n.d.)

Implementasi Item Based Collaborative Filtering

Metode *item-based collaborative filtering* menggunakan item yang disukai sebelumnya sebagai dasar perbandingan. Metode ini menentukan bagaimana sebuah objek yang telah diberi rating dibandingkan dengan objek lain dan kemudian mengidentifikasi item terkait yang memiliki kemiripan yang sesuai. Bobot nilai kemiripan yang disebutkan di atas kemudian digunakan untuk menentukan peringkat item. (Jaja et al., 2020)

Item based collaborative filtering menghitung kemiripan item berdasarkan atribut dari hasil survei kesukaan pengguna yang memiliki kesamaan dengan atribut dari dataset atau fitur yang serupa berdasarkan perhitungan *cosine similarity* kesamaan sepeda motor yang paling cocok akan direkomendasikan kepada pengguna.

Nilai *similarity* digunakan dalam pengukuran relevansi/kesamaan antara himpunan berdasarkan irisan kedua, selain itu juga nilai *similarity* dapat menentukan tingkat kesamaan antar titik data yang digunakan. Pengukuran *similarity* juga dapat menentukan pengujian kemiripan antar teks. Salah satu metode dalam menentukan nilai *similarity*

adalah *cosine similarity*, yang dapat menghitung kemiripan/hubungan antar kueri dan dokumen.(Al Rasyid & Ningsih, 2024)

Evaluasi

Pada tahapan ini dilakukan evaluasi sistem bertujuan untuk memperoleh hasil dari kualitas sistem saat merkomendasikan, dengan cara memisahkan data latih dan data uji yang bertujuan agar memudahkan saat dilakukan evaluasi pada sistem rekomendasi sepeda motor menggunakan metode *content based filtering* dan *item based collaborative filtering*. Pada tahapan evaluasi ini akan menggunakan metode perhitungan *precision* dan *recall* untuk menemukan hasil evaluasi yang relevan.

Tabel 1 Deskripsi Atribut Data

No	Atribut	Keterangan
1	Brand	Merk
2	Model	Model Sepeda Motor
3	Category	Tipe Sepeda motor
4	Displacement	ukuran mesin sepeda motor
5	Power	daya keluaran maksimum
6	Rating	ulasan rata-rata dari 5 bintang
7	Year	Tahun produksi
8	Engine Cylinder	jumlah silinder di mesin
9	Engine Stroke	jumlah tahapan untuk menyelesaikan satu langkah daya mesin
10	Gearbox	jumlah roda gigi dalam transmisi
11	Fuel Capacity	kapasitas maksimum tangki bahan bakar
12	Fuel System	sistem pengiriman bahan bakar ke dalam mesin
13	Cooling System	sistem pendingin mesin
14	Transmission type	jenis transmisi sepeda motor
15	Dry weight (kg)	berat sepeda motor
16	Wheelbase (mm)	jarak antara titik di mana roda depan dan belakang

17	Seat height (mm)	tinggi dari dasar tempat duduk ke tanah
18	Front brakes	jenis rem depan
19	Rear brakes	jenis rem belakang
20	Front tire	ukuran ban depan
21	Rear tire	ukuran ban belakang

C. Hasil dan Pembahasan

Sistem rekomendasi sepeda motor ini dibuat melalui berbagai tahapan yang telah direncanakan dalam metode. sistem ini berhasil diimplementasikan sesuai dengan tujuan yang diinginkan. Proses pembuatannya melibatkan beberapa langkah, mulai dari pengumpulan data, *preprocessing data*, hingga implementasi *content-based filtering* dan *item-based collaborative filtering*. Selain itu, antarmuka pengguna yang sederhana dan mudah dipahami juga dikembangkan, sehingga pengguna dapat dengan mudah mendapatkan rekomendasi sepeda motor yang sesuai dengan preferensi mereka. Evaluasi terhadap sistem ini dilakukan menggunakan metrik seperti *precision*, *recall*, dan *F1-score*, untuk menilai seberapa baik sistem ini dalam memberikan rekomendasi yang relevan.

Pengumpulan Data

Jumlah data sepeda motor yang didapatkan dari situs kaggle berjumlah 3966 data sepeda motor yang diambil dari beragam jenis *spesifikasi*.

Preprocessing Data

Tahapan *pre-processing* dimana pada tahapan ini adalah tahapan perubahan data Sepeda Motor yang akan dicari kemiripannya dengan data produk yang tersedia, tahapan pada *pre-processing* dibagi menjadi 2 proses, proses pertama *Data Cleaning* dan

Data Selection proses kedua *Tokenizing* dan *Stopword removal*

1. *Data cleaning* dan *data selection*

Pada tahap *data cleaning* yaitu dilakukan pembersihan data untuk penganalisaan yang tidak mempunyai data yang lengkap agar dapat dengan mudah mencari data. *Tahap data selection* yaitu dilakukan seleksi data-data yang sudah memiliki informasi lengkap untuk setiap *attribute* dan data tersebut dilakukan seleksi untuk mengelompokkan *attribute* sesuai dengan informasi yang dibutuhkan. (Mabrur, 2012)

2. *Tokenizing* dan *stopword removal*

Pada tahap kedua dari *preprocessing* adalah *tokenizing*, di mana pada langkah ini, terjadinya pemotongan string input berdasarkan setiap karakter yang tersusun. Langkah ini juga mencakup tokenisasi, yang mengekstrak teks berdasarkan karakter seperti spasi, tab, dan baris baru. Setelah tokenisasi selesai, langkah selanjutnya adalah *stopword removal*. Proses ini melibatkan penyortiran kata-kata penting dari hasil token (membuang kata-kata yang tidak terlalu penting) atau *wordlist* (simpan kata-kata penting). Hasil dari *preprocessing* menggunakan *tokenizing* dan *stopword removal* disimpan dalam dataset pada kolom *processed_text* yang dapat yang ditunjukkan pada tabel 2

Tabel 2 Hasil Proses Tokenizing Dan Stopword Removal

aeon cobra 50 atv 49.3 3.0 2.6 2012 single cylinder two-stroke automatic 5.0 carburettor air shaft drive 129.0 1050.0 800.0 expanding brake (drum brake) single disc 19/7-8 18/10-8

aeon crossland x4 400 atv 346.0 20.1 3.5 2012 single cylinder four-stroke automatic 14.0 carburettor liquid shaft drive 236.0 1230.0 850.0 double disc expanding brake (drum brake) 23/7-12 23/10-12

aeon urban 350i scooter 313.0 22.8 3.6 2012 single cylinder four-stroke automatic 13.5 injection liquid chain 177.0 1545.0 815.0 single disc . hydraulic single disc . hydraulic 120/70-16 140/70-15

ajp pr4 125 enduro enduro / offroad 124.0 12.5 3.3 2010 single cylinder four-stroke 5-speed 7.5 carburettor air chain 105.0 1410.0 910.0 single disc . 2 piston calliper single disc . 4 piston calliper 90/90-21 120/90-18

ajp pr7 adventure 650 enduro / offroad 659.7 48.0 3.7 2018 single cylinder four-stroke 6-speed 17.0 injection . delphi liquid chain 155.0 1540.0 920.0 single disc . 2-piston calipers single disc . single-piston caliper 90/90-21 140/80-18

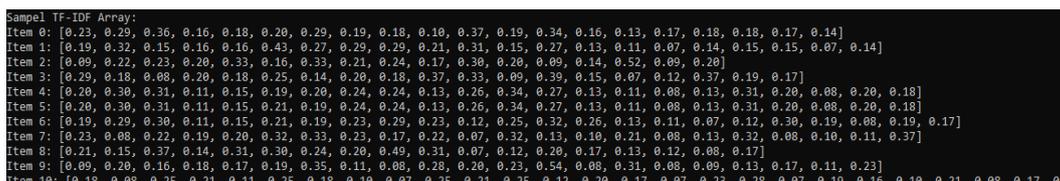
aprilia atlantic 250 scooter 250.0 22.8 3.7 2005 single cylinder four-stroke automatic 10.5 injection liquid belt 163.0 1470.0 678.0 single disc single disc 110/90-13 130/70-13

aprilia dorsoduro 1200 sport touring 1197.0 130.0 3.7 2015 v2 four-stroke 6-speed 15.0 injection liquid chain 212.0 1565.0 870.0 double disc . bremono . abs . single disc . bremono . abs . 120/70-zr17 190/55-zr17

aprilia dorsoduro 750 super motard 749.9 90.3 3.8 2011 v2 four-stroke 6-speed 12.0 injection . integrated engine management system . electronic fuel injection with ride-by-wire electronic throttle control . liquid chain 186.0 1505.0 870.0 double disc single disc 120/70-17 . 180/55-17 .

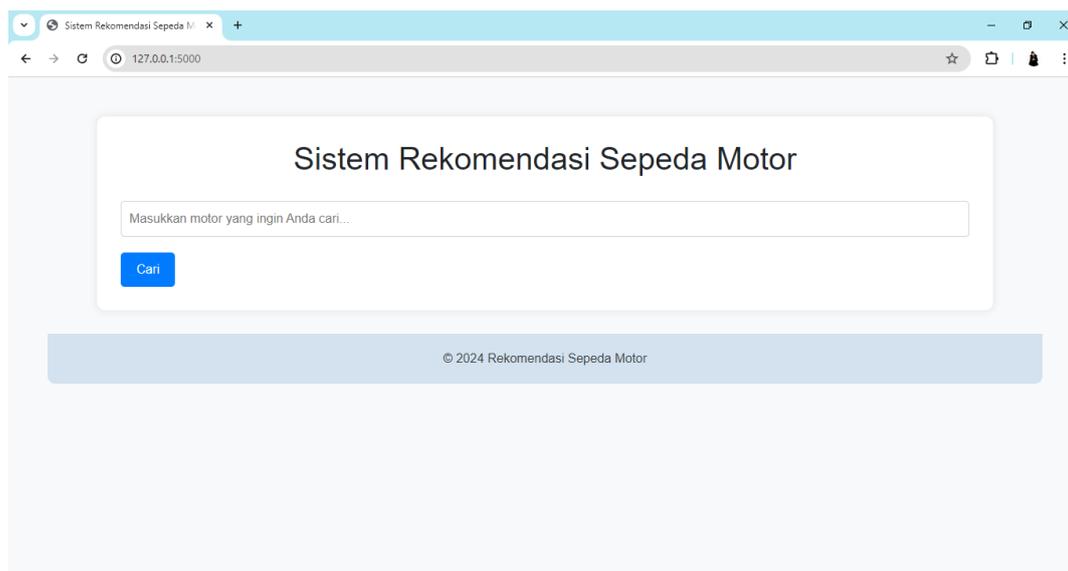
Implementasi Content Based Filtering

Metode *Content-based filtering* menghasilkan rekomendasi berdasarkan *query term* dari ketersediaan *content* yang terdapat pada *profile* sepeda motor. Variabel yang digunakan adalah *processed_text* yaitu hasil dari *preprocessing text* menggunakan *tokenizing* dan *stopword removal* berdasarkan dataset sepeda motor kemudian menghitung bobot TF-IDF. Perhitungan bobot TF-IDF dilakukan kepada setiap atribut atau informasi yang terdapat pada profil sepeda motor berdasarkan query. Hasil perhitungan bobot TF-IDF dapat dilihat pada gambar 1



Gambar 1 Hasil Perhitungan Tf-Idf

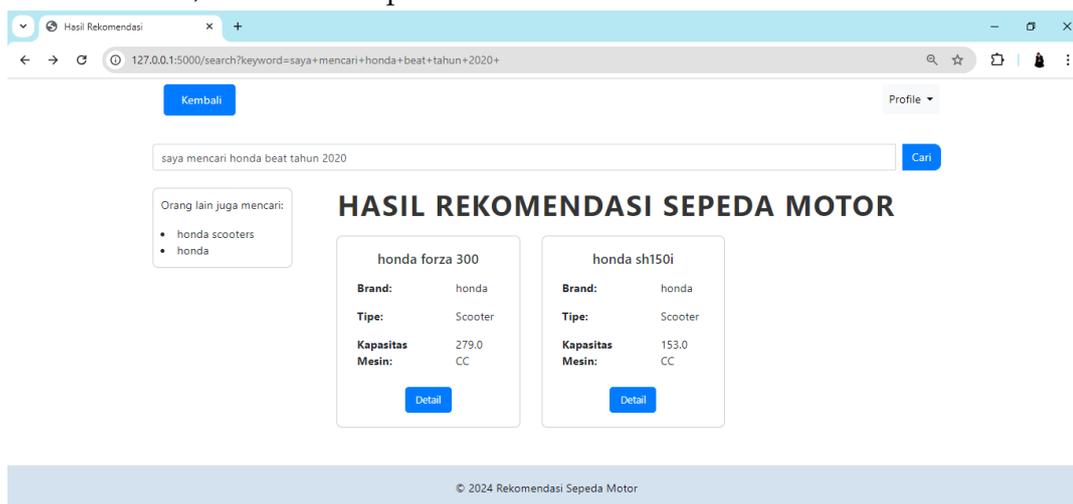
Dari proses perhitungan TF-IDF menghasilkan nilai bobot rekomendasi. Sepeda motor yang memiliki nilai bobot paling tinggi merupakan sepeda motor yang direkomendasikan, contoh hasil rekomendasi seperti yang ditunjukkan pada gambar 2



Gambar 5 Antarmuka Pencarian Sepeda Motor

Antarmuka Hasil Pencarian Sepeda Motor

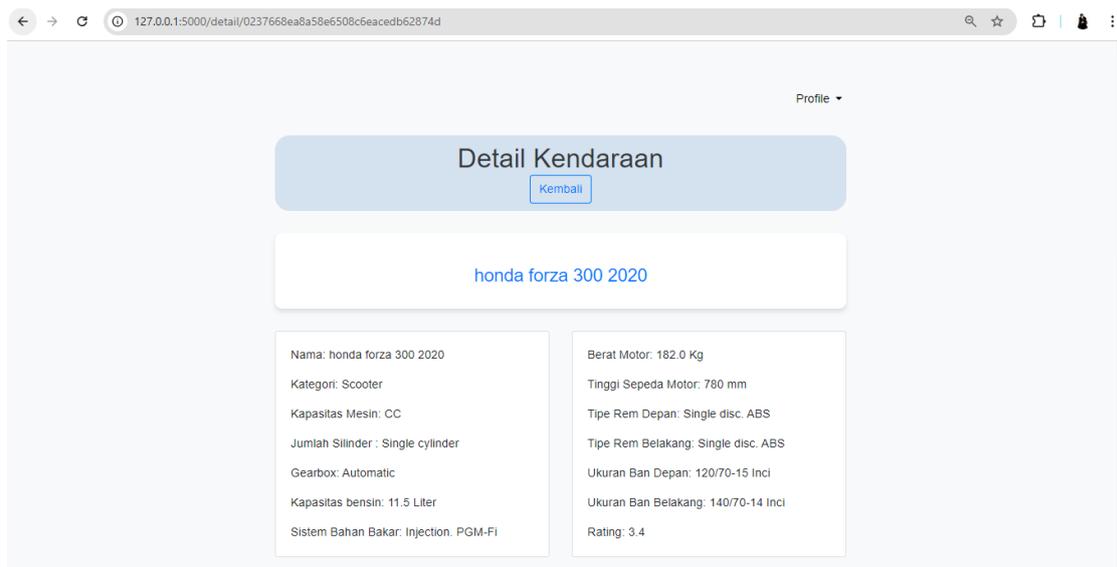
Pada Halaman hasil rekomendasi sepeda motor yang terdapat pada Gambar 6. Halaman ini terdapat beberapa hasil sepeda motor yang direkomendasikan sesuai dengan yang user cari pada halaman pencarian sepeda motor, lalu terdapat Form pencarian, button cari, button Kembali, rekomendasi pencarian dan button detail



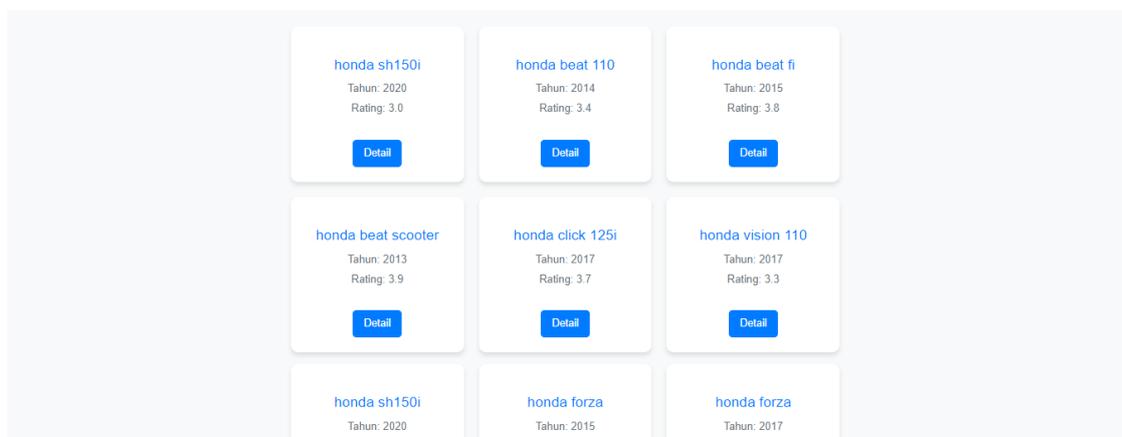
Gambar 6 Hasil Rekomendasi Sepeda Motor

Antarmuka Detail Rekomendasi Sepeda Motor

Pada Halaman Detail Rekomendasi sepeda motor yang dapat dilihat pada Gambar 7 dan Gambar 8. Terdapat hasil detail. lalu terdapat hasil rekomendasi lainnya yang terdiri dari sepeda motor yang mirip dengan sepeda motor pada detail.



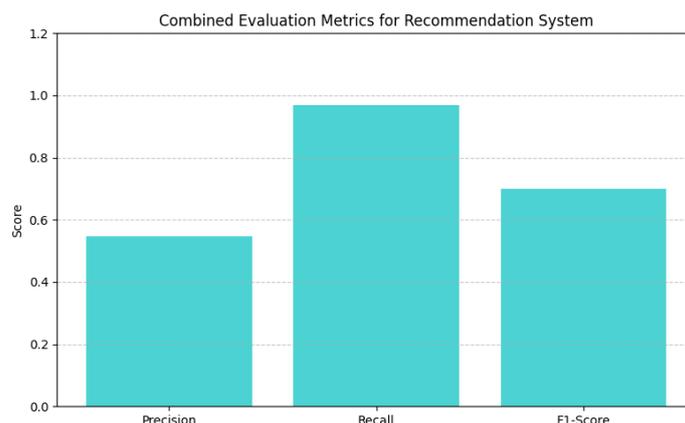
Gambar 7 Detail Rekomendasi



Gambar 8 Rekomendasi Lainnya

Evaluasi

Tahap ini melakukan evaluasi yang bisa digunakan pada data set yang bertujuan untuk memberikan hasil evaluasi untuk mengukur akurasi dari perangkat yang telah dibuat yaitu sistem rekomendasi. Pengujian dilakukan menggunakan *precision* untuk Mengukur seberapa banyak rekomendasi yang relevan dari semua rekomendasi yang diberikan, *recall* untuk Mengukur seberapa banyak item relevan yang direkomendasikan dari semua item relevan yang seharusnya direkomendasikan dan *f1-score*(Parwita, 2019)(Dewi et al., 2021) untuk menghitung nilai rata-rata dari *precision* dan *recall*. Penelitian ini menggunakan data uji dengan perbandingan 8:2 80% digunakan untuk data latih dan 20% digunakan untuk data uji. *True positif*(tp) yaitu menghitung berapa banyak item yang di rekomendasikan yang relevan lalu *false positif*(fp) yaitu menghitung berapa banyak item yang di rekomendasikan tidak relevan dan *false negative*(fn) yaitu menghitung berapa banyak item yang relevan namun tidak direkomendasikan.



Gambar 6 Hasil Evaluasi

Berdasarkan hasil *precesion* memberikan hasil 0.545 atau 54.5% lalu *recall* memberikan hasil evaluasi 0.970 atau 97,0% dan *f1-score* memberikan hasil 0.696 atau 69.6%

D. Kesimpulan

Penelitian ini berhasil membangun sistem rekomendasi sepeda motor menggunakan metode *content-based filtering* dan *item-based collaborative filtering*, dengan pemanfaatan TF-IDF untuk memberikan bobot pada setiap term dalam data kendaraan yang sudah dilakukan *preprocessing*. *Cosine similarity* digunakan untuk mengukur kesamaan antara item berdasarkan atribut yang dimiliki oleh sepeda motor. Evaluasi dilakukan menggunakan *precision*, *recall*, dan F1-score untuk menilai akurasi sistem. Hasil evaluasi menunjukkan bahwa sistem ini memiliki *precision* sebesar 54.5%, *recall* sebesar 97.0%, dan F1-score sebesar 69.6%.

Daftar Pustaka

- Al Rasyid, R., & Ningsih, D. H. U. (2024). Penerapan Algoritma TF-IDF dan Cosine Similarity untuk Query Pencarian Pada Dataset Destinasi Wisata. *Jurnal JTik (Jurnal Teknologi Informasi Dan Komunikasi)*, 8(1), 170–178. <https://doi.org/10.35870/jtik.v8i1.1416>
- Arfisko, H. H. (2022). Sistem Rekomendasi Film Menggunakan Metode Hybrid Collaborative Filtering Dan Content-Based Filtering. *E-Proceeding of Engineering*, 9(Vol.9, No.3(2022): Juni 2022), 2149. <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/18066>
- Dewi, I. P., Jondri, J., & Lhaksana, K. M. (2021). Prediksi Retweet Menggunakan Metode Bernoulli Dan Gaussian Naive Bayes Di Media Sosial Twitter Dengan Topik Vaksinasi Covid-19. *EProceedings of Engineering*, 8(5), 11216–11225.
- Fajriansyah, M., Adikara, P. P., & Widodo, A. W. (2021). Sistem Rekomendasi Film Menggunakan Metode Content Based Filtering. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 5(6), 2188–2199.
- Intakoris, S. dan S. S. (2015). Perkembangan Wilayah Pegunungan oleh Pengaruh Penggunaan Sepeda Motor. *Conference on Urban Studies and Development-1, Semarang, 8 September 2015*, 177–188. <http://proceeding.cousd.org>

- Jaja, V. L., Susanto, B., & Sasongko, L. R. (2020). Penerapan Metode Item-Based Collaborative Filtering Untuk Sistem Rekomendasi Data MovieLens. *D’CARTESIAN*, 9(2), 78. <https://doi.org/10.35799/dc.9.2.2020.28274>
- Mabrur, L. (2012). Penerapan Data Mining Untuk Memprediksi Program Studi Teknik Informatika Jurnal Komputer dan Informatika (KOMPUTA). *Jurnal Komputer Dan Informatika (KOMPUTA)*, 1, 53–57.
- Parwita, W. G. S. (2019). Pengujian Akurasi Sistem Rekomendasi Berbasis Content-Based Filtering. *Informatika Mulawarman: Jurnal Ilmiah Ilmu Komputer*, 14(1), 27. <https://doi.org/10.30872/jim.v14i1.1272>
- Penjualan, V., Motor, S., & Indonesia, E. (2024). *Penjualan Sepeda Motor Indonesia Naik pada Mei 2024*. 2023–2024.
- Raharjo, P. N., Handojo, A., & Juwiantho, H. (n.d.). *Sistem Rekomendasi Content Based Filtering Pekerjaan dan Tenaga Kerja Potensial menggunakan Cosine Similarity*.