

Penyusunan Sistem Rekomendasi Produk Diecast Mobil Dengan Metode Content-Based Filtering (CBF)

Anak Agung Aditya Nugraha^{a1}, Ngurah Agus Sanjaya ER^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹gungwah1927@gmail.com
²agus_sanjaya@unud.ac.id

Abstract

The growing popularity of diecast car collections has created a demand for efficient recommendation systems to assist collectors in discovering new products. This study focuses on the development of a content-based filtering (CBF) recommendation system for diecast car products. The system employs the TF-IDF (Term Frequency-Inverse Document Frequency) and cosine similarity techniques to calculate the relevance between products and user preferences. By analyzing the textual features of diecast car products, such as brand, model, and specifications, the CBF system generates personalized recommendations based on similarity scores. The evaluation of the system's performance demonstrates its effectiveness in providing accurate and relevant recommendations, which enhance the user experience and facilitate the exploration of the diecast car market.

Keywords: CBF, Diecast cars, Recommendation System, TF-IDF, Cosine Similarity

1. Pendahuluan

Dengan munculnya berbagai platform e-commerce serta pesatnya pertumbuhan industri diecast mobil, masyarakat semakin bergantung kepada teknologi informasi dengan tujuan memperoleh rekomendasi produk yang sesuai dengan preferensi mereka. Diecast mobil telah menjadi tren yang cukup populer di berbagai kalangan, khususnya kolektor serta penggemar otomotif dengan variasi model dan brand yang terus berkembang seiring zaman. Akan tetapi dengan semakin banyaknya pilihan yang tersedia, konsumen sering menghadapi kesulitan dalam mengelola berbagai opsi serta menemukan produk yang tepat untuk mereka. Kebutuhan akan sistem rekomendasi yang efektif dan efisien menjadi semakin meningkat. Sistem rekomendasi dapat membantu konsumen dalam menemukan produk diecast mobil yang sesuai dengan keinginan mereka serta mempermudah proses pemilihan dan pembelian. Dengan terciptanya sistem rekomendasi yang andal, konsumen dapat menghemat waktu serta usaha dalam mencari produk yang sesuai serta meningkatkan kepuasan mereka dengan hasil pembelian. Agar dapat meningkatkan akurasi sistem rekomendasi, pendekatan yang efektif dalam menganalisis konten produk yaitu menggunakan metode Content-Based Filtering (CBF). Metode CBF berfokus pada karakteristik atau atribut dari produk tersebut, seperti deskripsi, fitur, ataupun kategori. Maka dari itu, pada penelitian ini akan dilakukan penyusunan sistem rekomendasi produk diecast mobil menggunakan metode CBF yang akan dibantu oleh penggunaan Term Frequency-Inverse Document Frequency (TF-IDF) dan juga Cosine Similarity berdasarkan dataset deskripsi produk diecast mobil yang tersedia. Penggabungan metode TF-IDF dan Cosine Similarity ini bertujuan mengembangkan sistem rekomendasi yang dapat memberikan rekomendasi produk diecast mobil yang akurat dan juga relevan. Dengan adanya sistem rekomendasi ini, diharapkan konsumen dapat menemukan produk diecast mobil dengan mudah dan sesuai dengan preferensi, serta meningkatkan efisiensi dan kepuasan dalam proses pembelian.

2. Metode Penelitian

2.1 Sistem Rekomendasi

Sistem rekomendasi didefinisikan sebagai suatu sistem yang di mana bertugas memberikan saran/rekomendasi kepada pengguna sistem yang bersifat personal (tiap pengguna memiliki saran yang berbeda) [1]. Sistem ini dapat ditemukan dalam berbagai konteks, seperti e-commerce, layanan streaming musik, platform media sosial, dan lain sebagainya. Sistem rekomendasi memiliki tujuan dalam memahami preferensi pengguna serta menyajikan konten dan produk yang relevan. Pemahaman preferensi tersebut dilakukan dapat dengan pengumpulan data secara eksplisit (preferensi yang diberikan secara langsung) atau secara implisit (melacak perilaku saat menggunakan sistem). Sistem rekomendasi dibagi dalam beberapa klasifikasi berdasarkan metode rekomendasi yang sering digunakan, yaitu content-based recommendation, collaborative filtering recommendation, knowledge-based recommendation, serta hybrid recommendation [2].

2.2 Content-Based Filtering

Content Based Filtering (CBF) didefinisikan sebagai salah satu metode dalam sistem rekomendasi yang di mana menggunakan informasi mengenai atribut ataupun karakteristik item untuk memberikan rekomendasi kepada pengguna. Metode ini berfokus pada kesesuaian konten item dengan preferensi pengguna [3]. Content-Based Filtering memiliki keuntungan dalam kemampuannya untuk memberikan rekomendasi yang personal dan relevan berdasarkan preferensi pengguna. Sistem tidak bergantung pada data pengguna dan dapat bekerja dengan baik untuk pengguna baru yang belum memberikan banyak umpan balik.

2.3 Term Frequency-Inverse Document Frequency

Term Frequency-Inverse Document Frequency (TF-IDF) didefinisikan sebagai algoritma/metode yang digunakan dalam pemrosesan bahasa alami (NLP) dan juga sistem informasi dengan tujuan mengukur seberapa penting sebuah kata dalam sebuah dokumen atau korpus teks. TF-IDF berasal dari bidang information retrieval, namun pada saat ini banyak digunakan dalam perbandingan dokumen [4]. Metode ini menggabungkan dua buah komponen utama, yaitu frekuensi kata dalam dokumen (TF) dan invers dari frekuensi kata dalam korpus (IDF). Metode TF-IDF sering digunakan dalam berbagai aplikasi pemrosesan bahasa alami seperti klasifikasi teks, analisis sentimen, dan juga sistem rekomendasi berbasis teks. Rumus dari TF-IDF yaitu :

$$TF - IDF = TF * IDF \quad (1)$$

- a. Term Frequency (TF), didefinisikan sebagai pengukuran seberapa sering sebuah kata muncul pada dokumen. Salah satu rumus dari TF yaitu:

$$TF = \frac{\text{jumlah kemunculan kata dalam dokumen}}{\text{jumlah total kata dalam dokumen}} \quad (2)$$

- b. Inverse Document Frequency (IDF), didefinisikan sebagai pengukuran seberapa penting sebuah kata dengan memperhatikan seberapa umum atau jarang kata tersebut dalam keseluruhan korpus teks. Rumus yang digunakan dalam IDF yaitu:

$$IDF = \log \frac{\text{jumlah dokumen pada korpus}}{\text{jumlah dokumen yang mengandung kata}} \quad (3)$$

dalam rumus tersebut dihitung logaritma dari rasio jumlah total dokumen dalam korpus dengan jumlah dokumen yang mengandung kata tersebut yang dilakukan untuk memberikan bobot yang lebih besar pada kata yang jarang muncul dalam korpus, sehingga kata tersebut dianggap lebih penting.

2.4 Cosine Similarity

Cosine similarity didefinisikan sebagai metode yang digunakan dalam mengukur sejauh mana dua vektor berada dalam arah yang sama atau memiliki kesamaan dalam ruang vektor. Metode ini sering digunakan dalam pemrosesan bahasa alami serta sistem rekomendasi untuk membandingkan kemiripan antar dua dokumen atau dua vektor teks. Rumus cosine similarity mengukur sejauh mana dua vektor berada dalam arah yang sama atau memiliki kesamaan dalam ruang vektor. Nilai cosine similarity berkisar antara -1 hingga 1, dengan 1 menunjukkan arti kesamaan sempurna antar dua vektor, sedangkan 0 menunjukkan ketidaksamaan atau ketidakberhubungan antar dua vektor, dan -1 menunjukkan kebalikan atau keberlawanan antar dua vektor. Cosine similarity memiliki rumus sebagai berikut :

$$\cos \theta = \frac{(A \cdot B)}{\|A\| \|B\|} \quad (4)$$

- A dan B, merupakan vektor yang mewakili dua dokumen atau dua vektor teks yang ingin dibandingkan. Setiap komponen vektor mewakili nilai frekuensi atau bobot kata dalam dokumen.
- $(A \cdot B)$, merupakan hasil perkalian dot (dot product) antara vektor A dan vektor B (jumlah dari perkalian tiap komponen vektor yang sesuai).
- $\|A\| \|B\|$, merupakan panjang Euclidean (magnitude) dari masing-masing vektor A dan B yang dihitung dengan mengambil akar kuadrat dari jumlah kuadrat tiap komponen vektor.

3. Hasil dan Pembahasan

3.1. Pengumpulan Data

Data dikumpulkan dengan tujuan untuk memperoleh dataset yang mencakup informasi produk diecast mobil yang diperlukan dalam menyusun sistem rekomendasi. Dalam penelitian ini, data dikumpulkan menggunakan metode web scrapping dari website HobbyDB (<https://www.hobbydb.com/>) yang menyediakan berbagai list diecast dari berbagai model dan brand. Data yang diambil berupa link gambar, model, brand, skala, dan tahun rilis. Data tersebut disimpan dalam format CSV bernama hobbyDB.

3.2. Preprocessing Data

Setelah data dikumpulkan, dataset dipreprocessing dengan melalui beberapa proses, yaitu [5]:

- Case folding, yaitu mengubah data yang digunakan menjadi huruf kecil (lowercase) serta menghilangkan atau menghapus seluruh tanda baca yang terdapat pada data.
- Tokenizing, yaitu pemotongan string input dari data berdasarkan tiap kata yang menyusun data.
- Filtering (stopwords removal), yaitu menghilangkan atau menghapus kata yang kurang penting (tidak memiliki arti) pada data yang diproses.
- Stemming, yaitu mengubah kata menjadi kata dasar dengan menghilangkan imbuhan awalan, akhiran, ataupun kombinasinya.

3.3. Menghitung TF- IDF

Setelah dipreprocessing, dilakukan perhitungan TF-IDF. Perhitungan dilakukan dengan bobot TF-IDF untuk setiap kata dalam deskripsi produk berdasarkan frekuensi kemunculan kata dalam dataset serta seluruh dataset. Representasi tiap produk dapat dilakukan dengan berbagai bentuk, yaitu representasi matriks, representasi graf, representasi vektor, representasi model statistik, dan lain sebagainya.

3.4. Menghitung Cosine Similarity

Cosine similarity digunakan dalam tujuan mengukur tingkat kesamaan antara vektor representasi produk diecast mobil. Dengan membandingkan vektor produk diecast mobil tersebut dengan produk lainnya, maka akan dapat dihitung skor cosine similarity yang mencerminkan sejauh mana kedua produk memiliki kesamaan.

3.5. Pembentukan Sistem Rekomendasi

Dari hasil cosine similarity, pengguna dapat mempertimbangkan produk yang paling mirip dengan identifikasi produk diecast mobil yang memiliki skor tertinggi/ Dengan demikian, disusunlah sistem rekomendasi produk diecast mobil yang paling relevan dan sesuai dengan preferensi pengguna.

3.6. Evaluasi Sistem Rekomendasi

Dilakukan evaluasi kinerja sistem rekomendasi yang telah disusun. Evaluasi ini dilakukan dengan matriks evaluasi yaitu precision, recall, atau F1-score dengan tujuan mengukur sejauh mana sistem rekomendasi yang telah disusun memberikan rekomendasi yang akurat dan relevan.

4. Kesimpulan

Dalam penelitian ini telah disusun sebuah sistem rekomendasi produk diecast mobil yang menggunakan metode Content-Based Filtering (CBF) dengan pendekatan Term Frequency-Inverse Document Frequency (TF-IDF) dan Cosine Similarity. Sistem rekomendasi yang dikembangkan ini bertujuan untuk membantu konsumen dalam menemukan produk diecast mobil yang sesuai dengan preferensi mereka. Sistem rekomendasi ini diharapkan dapat memberikan kontribusi positif dalam memenuhi kebutuhan konsumen dan mendorong pertumbuhan industri diecast mobil.

Daftar Pustaka

- [1] N. Ula, C. Setianingsih, and R. A. Nugrahaeni, "SISTEM REKOMENDASI LAGU DENGAN METODE CONTENT-BASED FILTERING BERBASIS WEBSITE" e-Proceeding of Engineering, vol. 8, no. 6, pp. 12193–12199, 2021.
- [2] I. Diah, I. F. Rahmad, and A. Saleh, "IMPLEMENTASI RECOMMENDER SYSTEM PADA PEMILIHAN KAMERA MENGGUNAKAN CONTENT BASED DAN COLLABORATIVE FILTERING" IT (INFORMATIC TECHNIQUE) JOURNAL, vol. 7, no. 1, pp. 1-12, 2019.
- [3] Daniel, B. Muiyawan, and T. Sutrisno, "PEMBUATAN APLIKASI E-COMMERCE BERBASIS WEB DENGAN FITUR REKOMENDASI MENGGUNAKAN METODE CONTENT-BASED FILTERING" JURNAL ILMU KOMPUTER DAN SISTEM INFORMASI, vol. 10, no. 1, pp. 1-6, 2022.
- [4] A. R. Harischandra, M. F. A. Pratama, Felix, and A. P. Laia, "APLIKASI PENDUKUNG DESAIN INTERIOR DENGAN SISTEM REKOMENDASI BERDASARKAN NAMA BRAND PERABOT MENGGUNAKAN ALGORITMA CONTENT-BASED FILTERING BERBASIS WEB" Jurnal SIFO Mikroskil 23, vol. 23, no. 1, pp. 1-16, 2022.
- [5] M. Alkaff, H. Khatimi, and A. Eriady, "Sistem Rekomendasi Buku Menggunakan Weighted Tree Similarity dan Content Based Filtering" Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, vol. 20, no. 1, pp. 193-202, 2020.