

Sistem Rekomendasi Mobil Berdasarkan *Demographic* dan *Content-Based Filtering*

Herastia Maharani^{#1}, Ferry Alexander Gunawan^{#2}

^{#1}Departemen Sistem Informasi, Institut Teknologi Harapan Bangsa

^{#2}Departemen Teknik Informatika, Institut Teknologi Harapan Bangsa

Jalan Dipatiukur no.80-84, Bandung - Indonesia

¹herastia@ithb.ac.id

²ferryalexanderg@gmail.com

Abstrak — Memilih mobil yang sesuai dengan kebutuhan bukan merupakan suatu hal yang mudah. Banyaknya alternatif pilihan mobil yang tersedia dengan spesifikasi yang beragam menyebabkan calon pembeli sering merasa kebingungan dalam menentukan pilihan. Masing-masing orang pun bisa memiliki parameter dan prioritas yang berbeda dalam memilih mobil, sehingga tipe mobil yang cocok bagi seorang pembeli belum tentu cocok bagi pembeli yang lain. Penelitian ini mengembangkan sebuah sistem rekomendasi untuk membantu proses pemilihan mobil bagi calon pembeli dengan menggabungkan dua buah metode. Metode *demographic filtering* digunakan untuk memberikan rekomendasi berdasarkan kemiripan profil antar pembeli. Sedangkan metode *content-based filtering* digunakan untuk memberikan rekomendasi berdasarkan kemiripan antara kriteria mobil yang diinginkan pembeli dengan spesifikasi mobil yang tersedia. Hasil pengujian menunjukkan bahwa *content-based filtering* memberikan rekomendasi yang lebih baik dibandingkan dengan *demographic filtering*.

Kata kunci — mobil, sistem rekomendasi, *content-based filtering*, *demographic filtering*

Abstract — *Choosing a car that suits your need is not an easy task. The numbers of alternatives with various specifications often overwhelm and confuse the customer when they are trying to buy a car. Each customer possibly has different parameters and priorities in choosing a car, thus a car that fits one person's criteria might not be suitable for others. This study develops a recommendation system that can help the process of choosing a car by combining two methods. Demographic filtering is used to give recommendation based on the similarity between customer's profiles. Content-based filtering is used to give recommendation based on the similarity between customer's criteria and the specifications of available cars. Based on user evaluation, content-based filtering give better recommendations than demographic filtering.*

Keywords — car, recommendation system, *content-based filtering*, *demographic filtering*

I. PENDAHULUAN

Perkembangan teknologi seperti dengan adanya internet, Seiring dengan pesatnya perkembangan jaman dan teknologi saat ini, perkembangan dunia otomotif juga mengalami peningkatan yang cukup pesat. Sekarang hampir setiap orang menggunakan mobil, baik itu untuk memenuhi kebutuhan *user* maupun hanya sekedar memenuhi gaya hidup atau *lifestyle*. Banyak perusahaan-perusahaan yang bergerak dalam bidang industri mobil menawarkan mobil dengan teknologi canggih dan mengutamakan *safety first*. Tidak heran jika saat ini, mobil menjadi salah satu kendaraan penting dalam kehidupan sehari-hari. Dengan kata lain kebanyakan orang ingin memiliki mobil yang “cocok” bagi dirinya, baik dari segi teknologi yang digunakan maupun fitur-fitur yang ditawarkan. Secara tidak langsung, jika seseorang membeli mobil yang sesuai dengan kebutuhannya, maka spesifikasi dari mobil yang dibeli tersebut dapat menunjukkan preferensi dari si pembeli.

Sayangnya, mencari mobil yang “cocok” bagi seorang pembeli bukanlah hal yang mudah. Alternatif pilihan mobil yang semakin banyak ditambah dengan spesifikasi dan fitur yang beragam membuat banyak calon pembeli mengalami kesulitan dalam memilih mobil yang sesuai dengan kebutuhan mereka. Cara yang biasa digunakan untuk membantu pembeli memilih mobil adalah dengan datang ke suatu *showroom* mobil dan meminta pendapat dari orang yang lebih profesional dalam bidangnya. Namun tidak semua *showroom* mobil memberikan jawaban atau rekomendasi yang tepat karena setiap pembeli mempunyai karakteristik, kebutuhan dan pertimbangan harga yang berbeda-beda.

Untuk membantu memberikan solusi untuk masalah di atas, dalam penelitian ini dikembangkan sebuah sistem rekomendasi yang dapat memberikan alternatif mobil yang sesuai dengan kebutuhan pembeli berdasarkan karakteristik pembeli dan juga spesifikasi mobil yang menjadi preferensi pembeli. Sistem rekomendasi ini menggabungkan metode *Demographic Filtering* dan metode *Content-Based Filtering* dengan tujuan agar rekomendasi yang dihasilkan betul-betul sesuai dengan profil dan preferensi pembeli.

II. DEMOGRAPHIC FILTERING

Metode *demographic filtering* umumnya digunakan untuk

menangani *user* yang baru mendaftar atau *user* yang belum memiliki *profile* atau catatan transaksi atau aktivitas dalam sistem (*history*). Rekomendasi yang diberikan kepada *user* diperoleh berdasarkan nilai atribut yang dimiliki oleh *user* tersebut. Dalam penelitian ini, atribut atau data demografis yang dicatat untuk *user* adalah usia, jenis pekerjaan dan tingkat pendapatan *user*. Ketiga atribut tersebut digunakan untuk menghitung kemiripan antara satu *user* dengan *user* lainnya berdasarkan kemiripan data demografisnya. Seorang *user* atau calon pembeli kemudian akan mendapatkan rekomendasi mobil sesuai dengan jenis mobil yang dinilai baik oleh pembeli lain yang memiliki data demografis yang mirip dengan *user* tersebut.

Dalam penelitian ini, kemiripan antara *user* A dengan *user* B dihitung dengan menggunakan persamaan *cosine similarity* sebagai berikut [3]:

$$\cos(W_A, W_B) = \frac{W_A \cdot W_B}{\|W_A\| \|W_B\|} \quad (1)$$

Keterangan:

W_A = vektor *user* A

W_B = vektor *user* B

Vektor *user* dibentuk dari nilai ketiga atribut yang dipilih untuk data demografis, yaitu usia, jenis pekerjaan, dan tingkat pendapatan. Atribut usia dan pendapatan memiliki nilai yang kontinyu sehingga perlu dilakukan diskretisasi untuk membagi nilainya ke dalam beberapa interval. Klasifikasi untuk atribut usia dan pendapatan diberikan di Tabel 1 dan Tabel 2. Sedangkan untuk nilai atribut jenis pekerjaan dibagi menjadi 5 (lima) kelas yaitu Pelajar, Wiraswasta, Pegawai Negeri Sipil (PNS), Pegawai Swasta, dan Profesional. Sebagai contoh, seorang *user* dengan usia 35 tahun, bekerja sebagai wiraswasta, dan memiliki pendapatan Rp 7.000.000,- per bulan akan direpresentasikan dengan vektor (2, 2, 3).

Tabel 3 menunjukkan contoh profil beberapa *user* sesuai dengan klasifikasi atribut yang digunakan, beserta dengan ID mobil yang dimiliki atau dinilai baik. Misal terdapat calon pembeli A dengan representasi vektor (2, 1, 1). Untuk mendapatkan rekomendasi untuk *user* A maka dihitung nilai *cosine similarity* antara vektor *user* A dengan vektor *user* lain yang ada dalam sistem. Dengan perhitungan *cosine similarity* diperoleh nilai sebagai berikut:

$$\cos(W_A, W_{user-1}) = 0.69$$

$$\cos(W_A, W_{user-2}) = 0.89$$

$$\cos(W_A, W_{user-3}) = 0.95$$

$$\cos(W_A, W_{user-4}) = 1.0$$

Karena *user* A dan *user-04* memiliki kemiripan terbesar (100%), maka mobil yang direkomendasikan untuk *user* A adalah mobil yang dimiliki oleh *user-04*, yaitu mobil dengan ID CT-17 dan CT-18. *User* lain yang juga memiliki kemiripan yang tinggi dengan *user* A adalah *user-03* dengan nilai *cosine* 0.95, karena itu mobil dengan ID CT-08 juga dapat direkomendasikan ke *user* A namun dengan peringkat yang lebih rendah dibanding CT-17 dan CT-18.

TABEL 1.

KLASIFIKASI USIA

Usia	Kelas
$20 \leq \text{usia} < 30$	1
$30 \leq \text{usia} < 40$	2
$40 \leq \text{usia} < 50$	3
$50 \leq \text{usia} < 60$	4
$\text{Usia} \geq 60$	5

TABEL 2.

KLASIFIKASI PENDAPATAN [1]

Pendapatan/bulan (Rp)	Kelas
$\text{Pendapatan} < 2.000.000$	1
$2.000.000 \leq \text{Pendapatan} < 5.000.000$	2
$5.000.000 \leq \text{Pendapatan} < 10.000.000$	3
$10.000.000 \leq \text{Pendapatan} < 20.000.000$	4
$\text{Pendapatan} \geq 20.000.000$	5

TABEL 3.

CONTOH PROFIL *USER* DAN MOBIL YANG DIMILIKI

ID User	Umur	Pekerjaan	Pendapatan	ID Mobil
User-01	1	5	3	CT-01, CT-02, CT-03, CT-04
User-02	3	4	4	CT-12, CTK-66
User-03	2	1	2	CT-08
User-04	2	1	1	CT-17, CT-18

III. CONTENT-BASED FILTERING

Content-Based Filtering merekomendasikan *item* berdasarkan korelasi antara isi (*content*) dari sebuah *item* dengan preferensi *user* sebagai pembanding [4]. Dengan memberikan rating terhadap suatu produk maka dapat dibentuk profil dari *user* yang bersangkutan. Dari profil inilah sistem dapat menganalisa karakteristik dan minat *user* tersebut. Sistem akan memberikan rekomendasi kepada *user* dengan membandingkan karakteristik *item* yang ada dengan *item* yang sudah pernah di rating oleh *user* tersebut. Misalkan *user* A memberikan rating terhadap *item* Y, setelah dianalisa ternyata *item* X mempunyai atribut yang sama atau mirip dengan *item* Y, maka *item* X akan direkomendasikan kepada *user*. Asumsi dasar yang digunakan adalah jika *user* menyukai *item* Y maka dia juga pasti menyukai *item* yang serupa, misalnya *item* X.

Dalam penelitian ini, rekomendasi yang diberikan dari *content-based filtering* diperoleh berdasarkan dari kemiripan antara preferensi mobil *user* dengan spesifikasi mobil yang tersedia dalam system. Kriteria preferensi mobil *user* dapat diperoleh dari data mobil yang telah dimiliki oleh *user* atau dari data mobil yang pernah diberi rating tinggi oleh *user*. Jadi, jika *user* pernah memiliki atau memberikan rating tinggi untuk sebuah mobil X, maka spesifikasi mobil X tersebut akan disimpan sebagai profil *user* tersebut. Dari profil ini, akan

dicari mobil apa saja yang memiliki spesifikasi yang mirip dengan profil *user* tersebut. Parameter yang dipilih untuk merepresentasikan mobil terdiri dari tipe mobil, isi silinder (cc), jenis bahan bakar, ketersediaan *airbag*, tipe transmisi,, ketersediaan sensor parkir, dan harga mobil. Tabel 4 dan Tabel 5 menunjukkan klasifikasi untuk atribut harga dan isi silinder. Untuk atribut harga, pembagian interval dilakukan berdasarkan nilai kuartil dari sampel data mobil yang digunakan dalam penelitian ini. Nilai atribut jenis bahan bakar terdiri dari premium, pertamax, pertamax plus, dan solar. Tipe mobil dibagi berdasarkan kapasitas penumpang menjadi *city car* atau *family car*. Atribut tipe transmisi bernilai biner dengan *default* 1 untuk tansmisi *automatic* dan 0 untuk transmisi manual. Adapun atribut *airbag* dan sensor parkir masing-masing bernilai biner dengan *default* nilai 1 menunjukkan adanya fitur tersebut dan nilai 0 berarti fitur tersebut tidak tersedia. Dengan demikian, setiap mobil akan direpresentasikan sebagai vektor dengan 7 elemen, dimana setiap elemen merepresentasikan nilai untuk sebuah parameter.

Untuk menghitung nilai kemiripan antara profil *user* dengan mobil lain, maka terlebih dahulu dihitung jarak (*distance*) antara profil *user* dengan mobil lain yang ada dalam sistem dengan menggunakan persamaan yang diadaptasi dari [2] sebagai berikut:

$$Dis(profile, candidate) = \sum_{i=1}^n (profile_i - candidate_i)^2 \quad (2)$$

Dimana:

- n = jumlah elemen dalam vektor mobil
- $profile_i$ = nilai parameter ke- i dari vektor profil *user*
- $candidate_i$ = nilai parameter ke- i dari vektor mobil yang dihitung *distance*-nya

Setelah nilai *distance* diperoleh, maka nilai kemiripan atau *similarity* dihitung berdasarkan rumus berikut [2]:

$$Sim(profile, candidate) = \frac{1}{1 + Dis(profile, candidate)} \quad (3)$$

TABEL 4.

KLASIFIKASI HARGA MOBIL

Harga Mobil (Rp)	Kelas
< Rp 202.412.500,00	1
202.412.500 ≤ Harga < 241.450.000	2
241.450.000,00 ≤ Harga < 324.650.000	3
Harga ≥ 324.650.000	4

TABEL 5.

KLASIFIKASI BOBOT ISI SILINDER

Silinder (cc)	Bobot
cc ≤ 1500	5
1500 < cc ≤ 2000	4
2000 < cc ≤ 2500	3
2500 < cc ≤ 3000	2
cc > 3000	1

Sebagai contoh, *user* A pernah memberi *rating* tinggi untuk mobil CT30 yang spesifikasinya direpresentasikan dengan vektor (1, 5, 2, 1, 1, 1, 2), dan data mobil lain yang ada dalam sistem diberikan di Tabel 6.

Dari contoh di atas, sistem akan menghitung kemiripan antara setiap mobil dengan profil *user*. Contoh perhitungan nilai *distance* antara profil *user* (CT30) dengan CT01 diberikan sebagai berikut:

$$Dis(profile, CT01) = (1-1)^2 + (5-4)^2 + (2-1)^2 + (1-1)^2 + (1-0)^2 + (1-1)^2 + (2-2)^2 = 3$$

Perhitungan *similarity* untuk CT-01:

$$Sim(profile, CT01) = \frac{1}{1+3} = 0.25$$

Dengan cara yang sama, maka diperoleh nilai *similarity* untuk semua mobil yang ada dalam sistem sebagai berikut:

$$Sim(profile, CT02) = 0.33$$

$$Sim(profile, CT03) = 1.0$$

$$Sim(profile, CT04) = 0.5$$

Berdasarkan hasil perhitungan di atas dapat ditarik kesimpulan bahwa mobil yang direkomendasikan untuk *user* A adalah mobil dengan ID CT03 karena mobil inilah yang memiliki kemiripan tertinggi dengan mobil CT30 yang mewakili profil *user*.

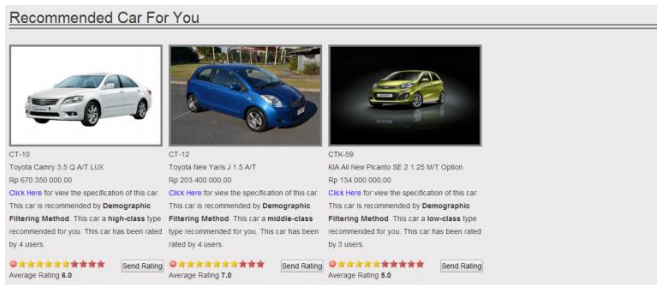
IV. HASIL PENGUJIAN DAN EVALUASI

Contoh hasil rekomendasi yang diperoleh dengan metode *demographic filtering* dan *content-based filtering* dapat dilihat di Gambar 1 dan 2. Untuk mengevaluasi hasil rekomendasi yang diberikan oleh sistem, dilakukan pengujian terhadap 20 orang *user* dengan latar belakang yang beragam. Setiap *user* diminta untuk memasukkan data demografinya dan juga spesifikasi mobil yang diinginkan atau jenis mobil yang disukai atau telah dimiliki. Kemudian, masing-masing *user* diminta untuk menilai hasil rekomendasi yang diberikan oleh sistem, baik rekomendasi yang dihasilkan melalui *demographic filtering* maupun yang dihasilkan melalui *content-based filtering*. Skala penilaian yang digunakan adalah 1-5 dengan penjelasan seperti pada Tabel 7. Adapun hasil evaluasi dari 20

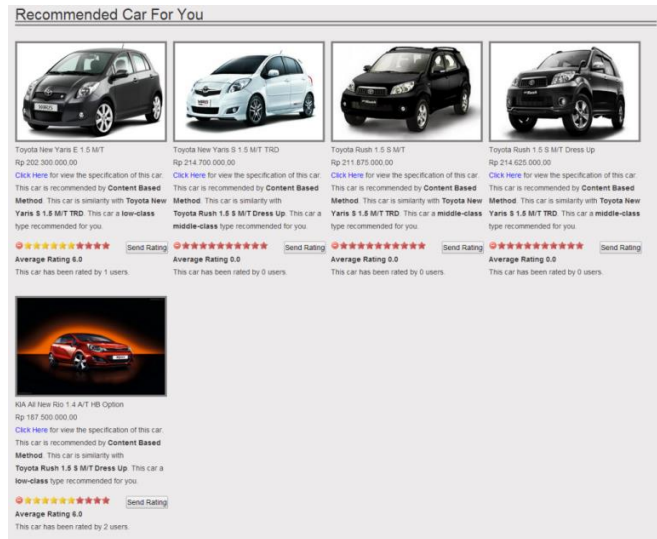
TABEL 6.

CONTOH DATA MOBIL YANG TERSEDIA

ID	Tipe mobil	CC	BBM	Airbag	Trans-misi	Sensor	Harga Mobil
CT01	1	4	1	1	0	1	2
CT02	1	5	1	1	1	1	1
CT03	1	5	2	1	1	1	2
CT04	1	5	2	1	1	1	3



Gambar 1. Contoh hasil rekomendasi dengan *Demographic Filtering*



Gambar 2. Contoh hasil rekomendasi dengan *Content-based Filtering*

TABEL 7.

KRITERIA PENILAIAN UNTUK EVALUASI *USER*

Nilai	Penjelasan
1	Tidak baik
2	Kurang baik
3	Cukup Baik
4	Baik
5	Sangat baik

TABEL 8.

HASIL PENILAIAN *USER* TERHADAP HASIL REKOMENDASI

User	Penilaian untuk hasil rekomendasi <i>Content-Based Filtering</i>	Penilaian untuk hasil rekomendasi <i>Demographic Filtering</i>
User A	5	4
User B	4	3
User C	3	4
User D	4	3
User E	4	3
User F	5	2
User G	3	2
User H	3	3
User I	3	3
User J	4	5
User K	4	4
User L	5	3
User M	3	3
User N	5	3
User O	5	4
User P	5	5
User Q	4	5
User R	4	3
User S	4	3
User T	3	3
Rata-Rata	4	3.4

V. KESIMPULAN

Berdasarkan data uji yang digunakan, metode *content-based filtering* menunjukkan hasil rekomendasi yang lebih baik dibandingkan dengan *demographic filtering*. Namun hal ini sangat dipengaruhi oleh pemilihan parameter yang digunakan untuk merepresentasikan profil *user* di masing-masing metode tersebut. Parameter yang digunakan untuk membangun profil *user* di metode *demographic filtering* dinilai masih kurang tepat untuk kasus pemilihan mobil, sehingga perlu dilakukan analisis lebih lanjut untuk menemukan data demografis apa yang lebih tepat digunakan untuk merepresentasikan profil *user*. Parameter yang digunakan dalam *content-based filtering* pun masih dapat dievaluasi lagi sehingga akurasi dari hasil rekomendasinya da-

pat lebih ditingkatkan. Selain pemilihan parameter, faktor lain yang berpengaruh terhadap hasil rekomendasi adalah cara mengklasifikasikan nilai setiap parameter yang digunakan. Teknik diskretisasi atau pengkodean yang berbeda akan sangat berpengaruh terhadap nilai kemiripan *user* ataupun *item* sehingga hasil rekomendasi pun dapat berubah.

REFERENSI

- [1] Direktorat Jendral Pajak Republik Indonesia. Pendaftaran NPWP Online. [Online]. Available: <http://ereg.pajak.go.id>
- [2] Meliana, Christianti and Hadiguna, Christian. *Aplikasi E-Commerce dengan Sistem Rekomendasi Berbasis Collaborative Filtering pada Toko Komputer Ekaria*. Jurnal Informatika, Universitas Kristen Maranatha. 2011.
- [3] Ricci, Francesco. *Recommender System Handbook*, London: Springer. 2011.

- [4] Van Meteren, Robin, Van Someren, Maartin. *Using Content Based Filtering for Recommendation*. [Online], Available: www.ics.forth.gr/~potamias/mlnia/paper_6.pdf.

Herastia Maharani menyelesaikan pendidikan sarjana di Departemen Teknik Informatika Institut Teknologi Bandung di tahun 2005 dan meraih gelar Magister Informatika di Institut Teknologi Bandung tahun 2010. Saat ini penulis bekerja sebagai staf pengajar di Institut Teknologi Harapan Bangsa, Bandung (ITHB). Bidang ilmu dan area penelitian yang ditekuni penulis adalah *data mining*, *information retrieval*, dan *social network analysis*.

Ferry Alexander Gunawan menyelesaikan pendidikan sarjana di Departemen Teknik Informatika Institut Teknologi Harapan Bangsa di tahun 2014.