



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB 2

LANDASAN TEORI

2.1 Sistem Rekomendasi

Sebuah rekomendasi adalah suatu saran atau anjuran yang membenarkan ataupun menguatkan sebuah objek [11]. Sistem rekomendasi merupakan suatu alat dan teknik yang mampu menetapkan saran atau anjuran untuk suatu objek yang dapat diminati oleh pengguna [12]. Untuk mengimplementasikan fungsi utamanya, sistem rekomendasi harus dapat memprediksi objek dilihat dari atributnya pendukungnya. Atribut tersebut yang mendukung suatu objek dapat direkomendasikan. Beberapa teknik untuk mengolah atribut tersebut menjadi rekomendasi adalah sebagai berikut.

2.1.1 Metode Sistem Rekomendasi

Berdasarkan pendekatan yang digunakan untuk menghasilkan rekomendasi, Ada lima teknik yang dapat digunakan dalam sistem rekomendasi [12].

1. Content-based System

Sistem ini akan merekomendasikan objek dihitung dari tingkat kesamaan suatu objek dengan objek lainnya berdasarkan fitur yang dimiliki masing-masing objek dengan preferensi pengguna. Misal seorang pengguna cenderung membeli mie ayam bawang, maka sistem akan memberikan rekomendasi produk-produk dengan rasa ayam bawang pada atributnya.

2. Collaborative Filtering System

Sistem ini akan merekomendasikan objek berdasarkan riwayat pilihan

pengguna lain yang tertarik dengan objek tersebut. Dari data pengguna tersebut sistem dapat memprediksi objek yang akan direkomendasikan ke pengguna lain yang memiliki ketertarikan yang sama.

3. Demographic System

Sistem ini akan merekomendasikan objek berdasarkan lokasi atau demografi penggunanya. Sistem ini dapat diasumsikan sebagai sistem yang dapat merekomendasikan objek yang sesuai dengan preferensi dari negara atau daerah tempat pengguna.

4. Knowledge-based System

Sistem ini menggunakan pengetahuan spesifik tentang bagaimana fitur-fitur suatu objek dapat memenuhi kebutuhan dan preferensi pengguna. Sebagai contoh, apabila pengguna dikategorikan sebagai *gamer* maka sistem akan merekomendasikan produk yang berspesifikasi tinggi. Sebaliknya, apabila pengguna dikategorikan sebagai *office-suite* maka sistem memberikan rekomendasi yang berspesifikasi menengah.

5. Hybrid System

Sistem ini merupakan gabungan antara teknik-teknik rekomendasi yang telah disebutkan. Sistem ini akan menggunakan lebih dari satu teknik untuk saling melengkapi kekurangan yang dimiliki teknik lainnya.

2.2 Content-Based Recommender System

Metode ini juga dapat disebut sebagai *Content-based Filtering*. Metode ini merekomendasikan produk menggunakan fitur atau atribut yang terdapat pada sebuah objek. Sekumpulan informasi yang ada mencerminkan karakteristik objek tersebut, seperti nama produk dan deskripsi produk. CBRS akan membandingkan

kemiripan data tersebut dengan data produk lainnya hingga ditemukannya nilai kemiripan antara satu produk dengan produk lainnya.

Kelebihan pada metode ini adalah sistem ini mampu memberikan rekomendasi tanpa membutuhkan personalisasi dari pengguna lainnya selama tersedianya informasi mengenai produk yang akan direkomendasi.

Disamping itu kekurangan dari metode ini yaitu *over-specialization* dalam arti hasil rekomendasi yang dihasilkan cenderung dapat diprediksi dan tidak berubah karena rekomendasi diberikan berdasarkan kemiripan dari produk yang dijadikan acuan [13].

2.3 Usaha Mikro, Kecil, Menengah

UMKM atau Usaha mikro, kecil, menengah merupakan kegiatan usaha produktif milik orang perorangan yang bertujuan untuk meningkatkan tingkat perekonomian nasional [14]. UMKM menjadi ujung tombak bagi pengusaha dalam menyalurkan dan menjual produk dari usaha besar ke konsumen. Pada prakteknya UMKM bukan sektor usaha yang bebas masalah. Perkembangan UMKM di Indonesia dapat berjalan dengan baik berdasarkan beberapa faktor dengan adanya pemodalan yang baik serta manajemen bisnis yang dilakukan secara tepat. Sebaliknya terbatasnya strategi pemasaran atau promosi terhadap bisnis yang dilakukan, dan lemahnya kreativitas dan inovasi. UMKM menjadi penyelamat ekonomi negara dalam keterpurukan. Terbukti pada data tahun 2012-2017 bahwa UMKM memiliki kontribusi dalam menyerap tenaga kerja sebesar 97,22% hal ini membantu meningkatkan perekonomian bangsa [14]. Dengan demikian UMKM mampu meningkatkan pendapatan masyarakat dan memiliki peran strategis dalam menekan angka kemiskinan.

2.4 Produk Sembako

Produk sembako atau sembilan bahan pokok merupakan kebutuhan pangan yang paling utama bagi manusia, karena untuk bertahan hidup manusia harus berusaha untuk melengkapinya dengan makanan atau minuman untuk dikonsumsi setiap hari. Menurut keputusan Kepmenperindag nomor 115 tahun 1998, kesembilan bahan itu adalah beras, minyak goreng, gula pasir, daging sapi dan daging ayam, telur ayam, susu, bawang merah dan bawang putih, gas dan minyak tanah, serta garam beryodium.

Untuk melengkapi kebutuhan tersebut umumnya manusia akan melakukan transaksi jual beli antara penjual dengan pembeli dalam proses itu juga pembeli dapat mencapai kesepakatan dengan penjual apabila barang yang dipilih telah sesuai dengan apa yang diinginkan pembeli.

Setiap produk memiliki varian dan manfaatnya masing-masing, sehingga nama, merk, ukuran, dan deskripsi produk merupakan hal yang penting dalam menentukan karakteristik tiap produk.

2.5 TF-IDF

Term frequency-inverse document frequency atau disingkat Tf-Idf merupakan teknik *text mining* yang digunakan dalam *content-based recommender system*. TF-IDF merupakan statistika numerik yang menghasilkan relevansi antara kata kunci dengan kumpulan dokumen lainnya, yang dengannya beberapa dokumen tertentu dapat diidentifikasi atau dikategorikan [15].

Term Frequency (TF) digunakan untuk mengukur seberapa banyak suatu *term* atau kata muncul pada dokumen. Dokumen berisi kumpulan kata yang mendukung suatu produk atau *item*. Dapat diketahui bahwa panjang total banyak

dokumen berbeda-beda, sehingga kemungkinan kata muncul lebih sering di dokumen yang besar daripada dokumen yang lebih kecil. Jadi untuk memperbaiki masalah ini, *term frequency* dinyatakan dengan formula sebagai berikut:

$$tf(t,d) = \frac{f_d(t)}{\max f_d(w)} \quad (2.1)$$

Keterangan:

$f_d(t)$: Jumlah term yang muncul pada dokumen.

$\max f_d(w)$: Jumlah kata pada dokumen.

Inverse Document Frequency (IDF) digunakan untuk mengukur seberapa penting sebuah term. Saat menghitung TF, semua term dianggap sama penting. Perlu diketahui bahwa kata-kata pendukung seperti "adalah", "dari", dan "itu" sering muncul dan dipertimbangkan tidak terlalu penting. Sehingga pembobotan term yang sering muncul dikurangi dan menaikkan term yang jarang, dengan formula sebagai berikut:

$$idf(t,D) = \log\left(\frac{N}{df_t}\right) \quad (2.2)$$

Keterangan:

N : jumlah seluruh dokumen

df_t : jumlah dokumen dengan term t dari seluruh dokumen

Dengan menggunakan *inverse document frequency*, maka nilai term yang sering muncul memiliki nilai *idf* rendah, dan term yang jarang muncul memiliki nilai *idf* tinggi. Sehingga formula pembobotan TF-IDF dapat dituliskan sebagai berikut:

$$w_{t,d} = tf_{t,d} \times \log\left(\frac{N}{df_t}\right) \quad (2.3)$$

Keterangan:

$w_{t,d}$: bobot untuk term t pada dokumen d

$tf_{t,d}$: jumlah kemunculan term t pada dokumen d

N : jumlah seluruh dokumen

df_t : jumlah dokumen dengan term t dari seluruh dokumen

2.6 Cosine Similarity

Cosine similarity merupakan salah satu metode untuk menghitung kesamaan dengan pendekatan geometris yang sering dipadukan dengan TF-IDF. *Cosine similarity* mengukur kemiripan antara dua vektor yang diperoleh dari perkalian sudut *cosinus* berdasar dua vektor yang dibandingkan [16].

Perhitungan kemiripan dilakukan antara dua dokumen yang telah dibentuk kedalam vektor (*tokenizing*) melalui metode TF-IDF. Representasi vektor digunakan untuk mempermudah perhitungan terhadap dokumen yang panjang [17]. Formula *cosine similarity* dapat dituliskan sebagai berikut:

$$similarity(A,B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}} \quad (2.4)$$

Keterangan:

$similarity(A,B)$: nilai *similarity* antara dokumen A dengan dokumen B

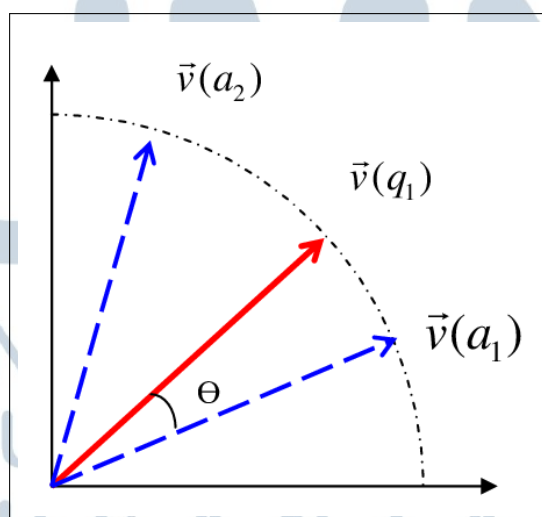
A_i : nilai bobot *term* i pada dokumen A

B_i : nilai bobot *term* i pada dokumen B

2.7 Vector Space Model

Mengukur tingkat kesamaan antara beberapa objek merupakan salah satu metode yang digunakan untuk beragam permasalahan, salah satunya *Natural Language Processing*. Dalam NLP mengukur similaritas antara teks banyak digunakan dalam mendeteksi plagiarisme, rekomendasi, hingga analisis sentimen [18].

Metode *Vector Space Model* merepresentasikan objek kedalam vektor. Dalam penelitian ini objek direpresentasikan sebagai vektor dari nilai fitur. Fitur tersebut mencirikan tiap objek dan mempunyai nilai numerik. Nilai dari tiap vektor didapatkan berdasarkan hasil bobot nilai *tf* (*term frequency*) dari tiap *term* pada sebuah dokumen. Seluruh *term* diperiksa pada tiap dokumen, apabila dokumen tersebut tidak mengandung sebuah *term*, maka bobot *term* pada dokumen tersebut bernilai nol.



Gambar 2.1. Ilustrasi *Vector Space Model*

Sumber: (Sidorov, et al)[19]

Pada gambar 2.1, tiap dokumen (data produk) dan *query* (keyword atau produk) yang direpresentasikan ke dalam vektor sesuai dengan panjang *query*. Vektor tersebut kemudian dihitung *cosinus* dari sudut yang terbentuk antara vektor *query* dan vektor dari tiap dokumen. Nilai yang terbentuk dari sudut tersebut merupakan nilai kesamaan atau *similarity* yang diukur menggunakan *Cosine Similarity*.

2.8 Precision dan Recall

Metode evaluasi yang digunakan pada penelitian ini adalah *precision and recall*. Metode *precision* merupakan kemampuan sebuah sistem dalam menggambarkan relevansi antara data yang diminta dengan hasil prediksi yang diberikan oleh model, sedangkan *recall* atau *sensitivity* menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi [20].

Untuk dapat mudah memahami metode evaluasi ini, perlu diketahui penggunaan *confusion matrix*. Dimana suatu kasus dapat terjadi empat kemungkinan, yaitu [21]:

- *True Positive*: Prediksi dikatakan positif dan itu benar. Contoh: Anda memprediksikan bahwa seorang wanita hamil dan wanita tsb memang benar hamil.
- *True Negative*: Prediksi dikatakan negatif dan itu benar. Contoh: Anda memprediksikan bahwa seorang pria tidak hamil dan benar ya pria kan tidak mungkin hamil.
- *False Positive*: Prediksi dikatakan positif dan itu salah. Contoh: Anda memprediksikan bahwa seorang pria hamil tetapi tidak mungkin pria bisa hamil.

- *False Negative*: Prediksi dikatakan negatif dan itu salah. Contoh: Anda memperkirakan bahwa seorang wanita tidak hamil tetapi sebenarnya wanita tersebut hamil.

Formula *precision* dan *recall* dinotasikan sebagai berikut [22]:

$$Recall = \frac{tp}{tp + fn} \quad (2.5)$$

$$Precision = \frac{tp}{tp + fp} \quad (2.6)$$

Uji coba akan dilakukan dengan menguji *string keyword* yang diterima oleh *searchbar* ke dalam sistem. Uji coba dilakukan beberapa kali sehingga mendapatkan rata-rata akurasi dari percobaan *precision* dan *recall*. Hasil akan diilustrasikan kedalam tabel sebagai berikut.

Tabel 2.1. Ilustrasi perhitungan *precision* dan *Recall*.

	<i>Relevant</i>	<i>Not Relevant</i>	Total
<i>Retrieved</i>	tp (hits)	fp (noise)	tp + fn
<i>Not Retrieved</i>	fn (miss)	tn (reject)	fn + tn
Total	tp + fn	fp + tn	tp + fn + fp + tn

Keterangan:

tp(*hits*): Dokumen relevan yang terpanggil

fp(*noise*): Dokumen tidak relevan yang terpanggil

fn(*miss*): Dokumen relevan yang tidak ditemukan

tn(*reject*): Dokumen tidak relevan yang tidak ditemukan

2.9 SUS (System Usability Scale)

System Usability Scale atau SUS adalah merupakan metode survey terdiri dari 10 buah pertanyaan yang pertama kali diperkenalkan oleh Brooke (1996) untuk mengukur kegunaan berbagai produk dan layanan. Meskipun banyak metode lainnya seperti *USE*, *CSUQ*, ataupun *PSSUQ*, SUS memiliki beberapa keunggulan signifikan. SUS dapat digunakan diberbagai produk layanan teknologi. Kedua, metode ini tersedia secara gratis, sehingga merupakan pilihan yang baik bagi para praktisi dan peneliti dalam hal ini [23].

SUS terdiri dari 10 buah pertanyaan. Responden diminta untuk memberi penilaian kegunaan produk menggunakan lima skala, dimulai dari satu (sangat tidak setuju) hingga lima (sangat setuju). Pertanyaan yang ditanyakan diadaptasi ke dalam Bahasa Indonesia oleh Sharfina dan Santoso (2016) dan telah divalidasi nilai korelasinya dengan versi aslinya [24].

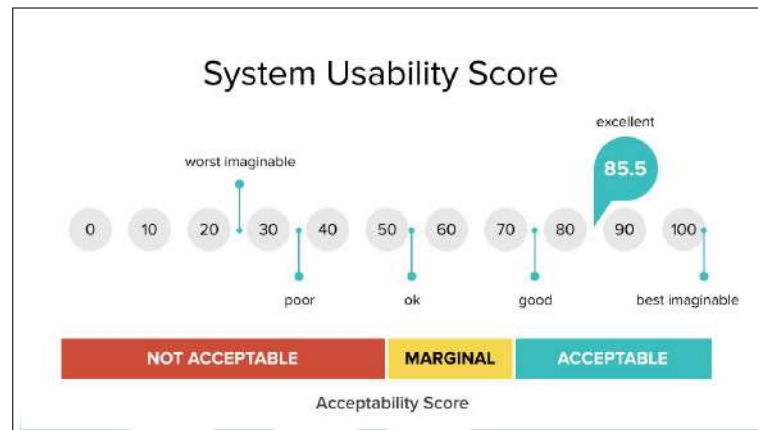


Tabel 2.2. Tabel pertanyaan SUS versi bahasa Indonesia

No.	Pertanyaan
1.	Saya berpikir akan menggunakan sistem ini lagi.
2.	Saya merasa sistem ini rumit untuk digunakan.
3.	Saya merasa sistem ini mudah untuk digunakan.
4.	Saya membutuhkan bantuan dari orang lain atau teknisi dalam menggunakan sistem ini.
5.	Saya merasa fitur-fitur sistem ini berjalan dengan semestinya.
6.	Saya merasa ada banyak hal yang tidak konsisten (tidak serasi) pada sistem ini.
7.	Saya merasa orang lain akan memahami cara menggunakan sistem ini dengan cepat.
8.	Saya merasa sistem ini membingungkan.
9.	Saya merasa tidak ada hambatan dalam menggunakan sistem ini.
10.	Saya perlu membiasakan diri terlebih dahulu sebelum menggunakan sistem ini.

Sumber: Sharfina & Santoso, 2016 [24]

Pengukuran dilakukan dengan menjumlahkan nilai dari setiap pertanyaan, nilai berkisar antara 0 sampai 4. Untuk pertanyaan positif pada nomor ganjil, nilai jawaban pengguna pada pertanyaan tersebut dikurangi 1 (satu), sementara pada pertanyaan negatif pada nomor genap, penilaiannya adalah 5 (lima) dikurangi dengan jawaban pertanyaan tersebut. Skor keseluruhan SUS didapat dengan menghitung total keseluruhan pertanyaan dikalikan dengan 2.5 untuk mendapatkan rentang nilai akhir 0-100 seperti yang digambarkan pada gambar.



Gambar 2.2. Rentang penilaian SUS

Sumber: (<https://medium.com/thinking-design/the-system-usability-scale-how-its-used-in-ux-b823045270b7>)

UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA