

# RECOMMENDATION SYSTEM WITH CONTENT-BASED FILTERING METHOD FOR CULINARY TOURISM IN MANGAN APPLICATION

Rhesa Havilah Mondy  
Informatika, Fakultas MIPA  
Universitas Sebelas Maret  
Jl. Ir. Sutami 36 A Surakarta  
rh.mondi@student.uns.ac.id

Ardhi Wijayanto  
Informatika, Fakultas MIPA  
Universitas Sebelas Maret  
Jl. Ir. Sutami 36 A Surakarta  
ardhi.wijayanto@staff.uns.ac.id

Winarno  
Informatika, Fakultas MIPA  
Universitas Sebelas Maret  
Jl. Ir. Sutami 36 A Surakarta  
win@staff.uns.ac.id

## ABSTRACT

*Culinary tourism has become one of the current trends in the culinary world. The amount of information circulating does not necessarily make culinary tourists easier to determine the desired menu choices. The use of search engines alone is still not enough, so we need a recommendation system that can provide advice according to user needs.*

*Content-based filtering method is able to produce recommendations that are user independence, making it suitable for use in developing culinary information providers such as the case studies of MANGAN applications where the number of users is still small and culinary data will always increase. This method recommends several objects based on the similarity of the selected object to the recommended object. Object similarity is calculated using the cosine similarity function based on profile items formed from a restaurant's content features.*

*Tests carried out on the results of recommendations with three different thresholds, obtained an average precision value of 0,8915 and an average accuracy value of 0,5118. The low value of accuracy is due to systematic errors. The results of this study are the content-based filtering method can be used to assist users in choosing restaurants based on the similarity of the item profile of a restaurant.*

**Keywords:** *Content Features, Item Profile, Cosine Similarity, Content-Based Filtering, Recommendation.*

## 1. PENDAHULUAN

Wisata kuliner menjadi salah satu tren masa kini di dalam dunia kuliner. Informasi terkait dunia kuliner dapat dengan mudah didapatkan terutama dalam bentuk media elektronik. Banyaknya informasi tersebut tidak serta merta membuat para wisatawan kuliner menjadi semakin mudah dalam menentukan pilihan menu hidangan yang diinginkan [1]. Mesin pencarian dapat membantu pengguna yang sudah memiliki referensi tentang hidangan apa yang hendak dikonsumsi. Kesulitan dihadapi bagi pengguna yang tidak memiliki referensi sama sekali. Penggunaan mesin pencarian saja dirasa masih belum cukup, sehingga diperlukan sebuah sistem rekomendasi yang dapat memberikan saran sesuai dengan kebutuhan pengguna [2].

Sistem rekomendasi memiliki dua pendekatan utama yaitu *collaborative filtering* dan *content-based filtering*. *Collaborative filtering* merupakan sistem rekomendasi yang memberikan rekomendasi item berdasarkan preferensi pengguna lain [2]. Metode *content-based filtering* merekomendasikan beberapa objek berdasarkan kemiripan

objek yang direkomendasikan dengan objek yang dipilih [3]. Metode ini tidak bergantung pada situasi apakah *item* tersebut merupakan *item* baru (yang belum pernah dipilih oleh pengguna manapun) maupun bukan *item* baru. Metode *content-based filtering* tidak menghasilkan rekomendasi yang tidak terduga. Seorang pengguna yang melihat menu ayam goreng, tidak menutup kemungkinan bahwa pengguna tersebut juga melihat menu daging babi. Hal-hal seperti inilah yang sebaiknya dihindari karena biasanya orang yang menyukai suatu *item* tertentu cenderung akan menyukai *item* lain dengan fitur serupa [4], maka dari itu dalam penelitian ini digunakan metode *content-based filtering*.

Penelitian serupa telah dilakukan oleh Maharani dan Gunawan [5] dengan membandingkan dua metode yang digunakan untuk sistem rekomendasi. Parameter pembentukan *user profile* yang digunakan masih belum tepat, tetapi hasil penelitian tetap menunjukkan bahwa metode *content-based filtering* lebih unggul daripada *demographic*. Penelitian lain juga dilakukan oleh Huda [6] dengan membentuk klasifikasi fitur konten berdasarkan karakter rasa pada makanan sebagai dasar penentu rekomendasi. Selera pengguna digambarkan dengan sembilan macam atribut, yakni *sweet, savory, salty, sour, bitter, sauce, spicy, vegetable* dan *meat*. *Pearson-correlation* digunakan sebagai metode perhitungan similaritas fitur konten antar makanan.

Sistem rekomendasi dengan metode *content-based filtering* mampu menghasilkan rekomendasi yang bersifat *user independence*, sehingga cocok digunakan untuk penyedia informasi kuliner yang sedang berkembang seperti pada studi kasus aplikasi MANGAN karena pengguna masih sedikit dan data kuliner akan selalu bertambah. Hasil dari penelitian ini adalah diberikan beberapa rekomendasi yang memuat gambar, nama, dan jarak dari restoran yang serupa berdasarkan kemiripan fitur konten ketika pengguna memilih restoran tertentu.

## 2. TINJAUAN PUSTAKA

### 2.1 Pengertian Sistem Rekomendasi

Sistem rekomendasi adalah peralatan perangkat lunak dan suatu teknik yang menyediakan rekomendasi *item* yang berguna untuk pengguna [3]. Rekomendasi diberikan berdasarkan berbagai macam proses pengambilan keputusan seperti *item* apa yang sebaiknya dibeli, musik apa yang sebaiknya didengar, makanan apa yang sebaiknya dikonsumsi atau berita *online* apa yang sebaiknya dibaca.

Sistem rekomendasi akan memberikan rekomendasi secara personal kepada setiap pengguna, bukan sekadar

memberikan daftar *item* yang paling diminati. Rekomendasi diberikan kepada pengguna terkait *item-item* yang mungkin sesuai dengan ketertarikan pengguna. Terdapat dua pendekatan utama dalam mengembangkan sistem rekomendasi yaitu *content-based filtering* dan *collaborative filtering*.

## 2.2 Content Based-Filtering

Sistem rekomendasi dengan metode *content-based filtering* merekomendasikan *item* yang mirip dengan *item* sebelumnya yang disukai atau dipilih oleh pengguna. Kemiripan *item* dihitung berdasarkan pada fitur-fitur yang ada pada *item* yang dibandingkan [3]. Metode ini bersifat *user independence*, tidak bergantung pada situasi apakah *item* tersebut merupakan *item* baru (yang belum pernah dipilih oleh pengguna manapun) maupun bukan *item* baru. Jika seorang pengguna telah memesan suatu menu hidangan pada kategori tertentu maka sistem akan mencoba merekomendasikan menu hidangan dengan kategori serupa yang juga tersedia di restoran lain yang mungkin akan disukai juga oleh pengguna tersebut.

Kelemahan dari metode *content-based filtering* adalah terbatasnya rekomendasi hanya pada *item-item* yang mirip sehingga tidak ada kesempatan untuk mendapatkan *item* yang tidak terduga. Akan tetapi, Pazzani dan Billsus dalam jurnalnya menjelaskan memang benar bahwa tidak menutup kemungkinan pengguna yang melihat atau membeli boneka juga telah membeli film dewasa, oleh sebab itu hal-hal seperti ini sebaiknya tidak direkomendasikan [4].

## 2.3 Preprocessing

*Preprocessing* adalah proses pengubahan bentuk data yang tidak terstruktur menjadi data yang terstruktur agar sesuai kebutuhan untuk diproses selanjutnya. Data yang telah melewati *preprocessing* adalah data yang lebih terstruktur [7]. Ada beberapa tahap *preprocessing* yang dilakukan yakni *case folding*, tokenisasi, eliminasi, *filtering* dan *stemming*. Penjelasan mengenai tahap-tahap tersebut adalah sebagai berikut:

1. *Case Folding* adalah tahap mengubah semua huruf dalam dokumen menjadi huruf kecil [8].
2. Tokenisasi adalah adalah tahap pemecahan kalimat berdasarkan tiap kata yang menyusunnya [8].
3. Eliminasi diterapkan untuk menghapus duplikasi kata. Kata yang duplikat diasumsikan memiliki fitur yang sama, sehingga jika dalam satu array terdapat beberapa data yang sama, hanya akan disimpan satu saja [9].
4. *Filtering* adalah tahap mengambil kata-kata penting dari hasil tahap sebelumnya. *Filtering* dapat dilakukan dengan menghilangkan *stoplist/stopword* (kata-kata yang tidak deskriptif, seperti kata “yang” dan “dari”) atau kata yang termasuk dalam *wordlist* [7].
5. *Stemming* adalah tahap transformasi suatu kata menjadi kata dasarnya (*root word*) [8].

## 2.4 Cosine Similarity

*Cosine similarity* merupakan perhitungan yang sering digunakan untuk menghitung kemiripan diantara *item-item* [10]. Secara umum, fungsi *similarity* adalah fungsi yang menerima dua buah obyek berupa bilangan riil (0 dan 1) dan mengembalikan nilai kemiripan (*similarity*) antara kedua obyek tersebut berupa bilangan riil.

*Cosine similarity* merupakan salah satu metode pengukuran kemiripan yang populer. Metode ini digunakan untuk

menghitung nilai kosinus sudut antara dua vektor dan biasanya digunakan untuk mengukur kemiripan antara dua dokumen. Fungsi *cosine similarity* antara *item* A dan *item* B ditunjukkan sebagai berikut [11].

$$\text{sim}(A, B) = \frac{n(A \cap B)}{\sqrt{n(A)n(B)}} \quad (1)$$

Keterangan:

$\text{sim}(A, B)$  = nilai similaritas dari *item* A dan *item* B  
 $n(A)$  = banyaknya fitur konten *item* A  
 $n(B)$  = banyaknya fitur konten *item* B  
 $n(A \cap B)$  = banyaknya fitur konten yang terdapat pada *item* A dan juga terdapat pada *item* B

Jika kedua objek memiliki nilai similaritas 1, maka kedua objek dikatakan identik dan sebaliknya. Semakin besar hasil dari fungsi *similarity*, maka kedua objek yang dievaluasi dianggap semakin mirip dan sebaliknya.

## 2.5 Top-N Recommendation

Kandidat *item-item* yang didapat dari perhitungan similaritas digunakan sebagai rekomendasi untuk pengguna. *Item-item* tersebut diseleksi menjadi hanya  $N$  *item* dengan nilai similaritas yang paling tinggi. *Item* serupa yang memiliki nilai similaritas tinggi diasumsikan cenderung lebih dipilih oleh pengguna [1].  $N$  *item* tersebut kemudian ditampilkan sebagai daftar *item* yang direkomendasikan untuk pengguna.

## 2.6 Confusion Matrix

*Confusion matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu sistem. Pada dasarnya *confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya [12]. Metode yang digunakan untuk mengukur performa sistem adalah presisi dan akurasi. Gagasan ini dapat dijelaskan dengan Tabel 1. Presisi adalah ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Akurasi didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual [13]. Presisi dan akurasi dirumuskan dengan Persamaan (2) dan Persamaan (3).

Tabel 1 *Confusion Matrix*

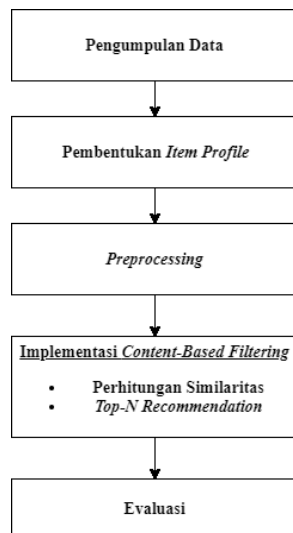
		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	TP (True Positive)	FP (False Positive)
	FALSE	FN (False Negative)	TN (True Negative)

$$\text{presisi} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

## 3. METODOLOGI

Tahap-tahap yang dilakukan dalam proses pengembangan sistem rekomendasi untuk wisata kuliner pada aplikasi MANGAN ini dapat dilihat pada Gambar 1.



Gambar 1 Metodologi Penelitian

### 3.1 Pengumpulan Data

Tahap awal yang dilakukan adalah pengumpulan data. Pengumpulan data dilakukan dengan cara mengambil data dari tabel keywordResto yang tersimpan pada basis data aplikasi MANGAN sejak Oktober 2016 hingga Maret 2018. Data yang didapat disimpan dalam file berekstensi .json. Tabel keywordResto terdiri dari dua kolom yakni index dan keyword. Kolom index berisi *id unique* suatu restoran, sedangkan kolom keyword berupa kalimat berisi fitur konten suatu restoran yang terdiri dari beberapa kata kunci meliputi nama restoran, nama menu, rasa, fasilitas, kategori, dan lokasi.

### 3.2 Pembentukan Item Profile

*Item Profile* selanjutnya dibentuk dari data yang telah dikumpulkan pada tabel keywordResto. *Item profile* terdiri dari kolom nama restoran dan kolom fitur konten. Data pada fitur konten diambil dari kolom keyword yang terdapat pada tabel keywordResto. *Item profile* dibentuk karena digunakan sebagai konten dari tiap restoran yang akan dibandingkan kemiripannya.

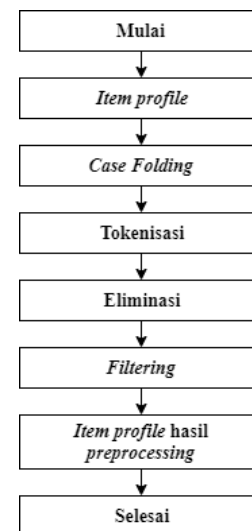
### 3.3 Preprocessing

*Item profile* yang telah dibentuk dari data yang telah dikumpulkan sebelumnya masih belum sesuai, maka dari itu tahap *preprocessing* perlu dilakukan untuk membuat *item profile* menjadi lebih terstruktur untuk proses selanjutnya. *Preprocessing* yang dilakukan dalam penelitian ini antara lain *case folding*, tokenisasi, eliminasi, dan *filtering*. Tahap *stemming* tidak dilakukan karena tidak ditemukan satupun kata kerja yang perlu diubah menjadi kata dasar pada data yang telah dikumpulkan. *Flowchart* tahap-tahap *preprocessing* ditunjukkan pada Gambar 2.

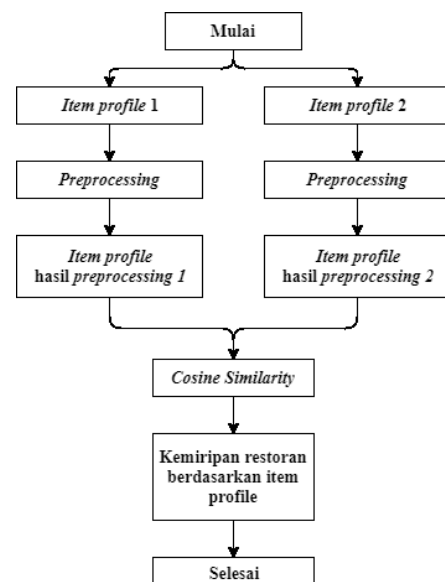
### 3.4 Implementasi Content-Based Filtering

Metode *content-based filtering* memiliki prinsip memberikan rekomendasi berdasarkan kemiripan *item profile*. Ketika pengguna memilih suatu restoran, maka akan diberikan rekomendasi berupa daftar restoran yang memiliki *item profile* yang mirip dengan restoran yang dipilih. *Item profile* dihitung kemiripannya dengan cara membandingkan *item profile* dari restoran yang dipilih oleh pengguna dengan *item profile* dari restoran lainnya setelah melalui tahap

*preprocessing*. *Flowchart* implementasi metode *content-based filtering* ditunjukkan pada Gambar 3.



Gambar 2 Flowchart Preprocessing



Gambar 3 Flowchart Implementasi content-based filtering

#### 3.4.1 Perhitungan Similaritas

Metode yang digunakan untuk menghitung kemiripan antar item profile adalah fungsi cosine similarity. Kedua restoran dikatakan identik jika perbandingan kedua item profile memiliki nilai similarity 1. Kedua restoran dikatakan berbeda jika kedua item profile memiliki nilai similarity 0. Semakin besar hasil dari fungsi cosine similarity kedua restoran yang dievaluasi dianggap semakin mirip dan sebaliknya.

#### 3.4.2 Top-N Recommendation

Hasil dari implementasi berupa nilai *cosine /* similaritas antara restoran yang satu dengan restoran lainnya. Kandidat *item-item* yang didapat dari perhitungan similaritas selanjutnya dijadikan rekomendasi untuk pengguna. *Item-item* tersebut diseleksi menjadi hanya lima *item* yang masuk kedalam *top-N recommendation* dengan nilai similaritas yang paling tinggi karena semakin tinggi nilai similaritas,

maka diasumsikan kedua restoran yang dibandingkan akan semakin mirip. Lima *item* tersebut kemudian ditampilkan sebagai daftar *item* yang direkomendasikan untuk pengguna ketika memilih sebuah restoran.

### 3.5 Evaluasi

Kualitas hasil rekomendasi selanjutnya diuji dengan cara dihitung presisi dan akurasinya. Pengujian dilakukan dengan cara membandingkan hasil rekomendasi dengan salah satu kategori dari restoran yang dipilih. Jika restoran hasil rekomendasi termasuk ke dalam kategori yang sama dengan restoran yang dipilih, maka hasil rekomendasi tepat sasaran dan sebaliknya.

Pengujian dilakukan sejumlah tiga kali dengan menggunakan parameter *threshold* yang berbeda untuk masing-masing pengujian. *Threshold* digunakan untuk membatasi jumlah rekomendasi restoran yang ditampilkan berdasarkan nilai similaritasnya. Nilai *threshold* ditentukan berdasarkan percobaan yang telah dilakukan sebelumnya. Jika *threshold* terlalu rendah, maka restoran yang muncul sebagai rekomendasi terlalu banyak, begitu pula jika terlalu tinggi, maka restoran yang direkomendasikan terlalu sedikit. Restoran dengan nilai similaritas dibawah nilai *threshold* tidak akan ditampilkan sebagai rekomendasi.

## 4. PEMBAHASAN

### 4.1 Pengumpulan Data

Data yang dikumpulkan merupakan kata-kata kunci untuk setiap restoran yang terdapat pada tabel keywordResto pada basis data aplikasi MANGAN. Data pada tabel keywordResto yang diambil adalah data restoran yang terdaftar sejak Oktober 2016 hingga Maret 2018. Restoran yang terdaftar di aplikasi MANGAN sebanyak 353 restoran. Menu yang terdaftar ada 2599 menu.

Tabel keywordResto terdiri dari dua kolom yakni index dan keyword. Kolom index berisi *id unique* suatu restoran, sedangkan kolom keyword berupa kalimat berisi fitur konten suatu restoran yang terdiri dari beberapa kata kunci meliputi nama restoran, nama menu, rasa, fasilitas, kategori, dan lokasi. Contoh tabel keywordResto direpresentasikan seperti pada Tabel 2. Kolom index berjumlah 353 indeks restoran dan kolom keyword berjumlah 66.724 kata kunci.

Tabel 2 Contoh keywordResto

index	keyword
ayamkeprabon	ayam geprek keprabon ayam kremes cappuccino chocolate choco cookies paket cup ayam geprek blenger paket cup ayam geprek jumbo paket geprek blenger paket geprek jumbo taro smoothies manis pedas keju mozarella ayam kremes halal delivery toilet ac khas indonesia keprabon

### 4.2 Pembentukan *Item Profile*

*Item Profile* selanjutnya dibentuk dari data pada tabel keywordResto. *Item profile* terdiri dari kolom nama restoran dan kolom fitur konten. Data pada fitur konten diambil dari kolom keyword yang terdapat pada tabel keywordResto.

Contoh kasus dikemukakan untuk mempermudah memahami proses pembangkitan rekomendasi dengan metode *content-based filtering*. Ketika pengguna memilih restoran Ayam Geprek Keprabon, maka akan dibandingkan *item profile* dari restoran tersebut dengan *item profile* dari

restoran lainnya untuk dicari kemiripannya. Restoran Soto Ayam Kremes As Ri dan restoran La Moda Del Gelato digunakan sebagai contoh kasus. *Item profile* yang terbentuk untuk masing-masing restoran ditampilkan pada Tabel 3.

Tabel 3 *Item Profile*

Nama Restoran	Fitur Konten
Ayam Geprek Keprabon	ayam geprek keprabon ayam kremes cappuccino chocolate choco cookies paket cup ayam geprek blenger paket cup ayam geprek jumbo paket geprek blenger paket geprek jumbo taro smoothies manis pedas keju mozarella ayam kremes halal delivery toilet ac khas indonesia keprabon
Soto Ayam Kremes As Ri	soto ayam kremes as ri es buah susu keju kremes paket b soto ayam komplit manis asin halal delivery toilet ac tv khas indonesia kartasura
La Moda Del Gelato	la moda del gelato chocolate muffin gelato ice cream red velvet cake quiche lorraine khas indonesia

### 4.3 Preprocessing

*Preprocessing* dilakukan untuk mengubah *item profile* yang telah dikumpulkan dari tahapan sebelumnya agar menjadi lebih sesuai, sehingga dapat digunakan untuk proses selanjutnya. Tahapan-tahapan yang dilakukan dalam *preprocessing* adalah *case folding*, tokenisasi, eliminasi, dan *filtering*.

#### 4.3.1 Case Folding

*Case folding* dilakukan sebagai tahapan awal dalam proses *preprocessing*. Kalimat dalam kolom fitur konten pada Tabel 3 dibuat menjadi huruf kecil semua agar data yang digunakan menjadi lebih sesuai. Pada contoh kasus yang digunakan tidak ditemukan satupun huruf kapital, sehingga proses dapat langsung dilanjutkan ke tahapan *preprocessing* selanjutnya yaitu tokenisasi.

#### 4.3.2 Tokenisasi

*Item profile* selanjutnya disesuaikan dengan tahap tokenisasi. Tokenisasi dilakukan dengan cara memisahkan kalimat menjadi bentuk tiap kata yang menyusunnya. Kalimat pada fitur konten dalam *item profile* pada Tabel 3 dipisahkan menjadi per kata. Hasil tokenisasi tersaji pada Tabel 4.

#### 4.3.3 Eliminasi

*Item profile* yang telah ditokenisasi masih belum sesuai. Contohnya pada Tabel 4 masih ditemukan beberapa kata yang masih serupa. Restoran Ayam Geprek Keprabon memiliki enam kata yang serupa antara lain: ayam, geprek, keprabon, kremes, paket, dan cup. Kata yang serupa pada restoran Soto Ayam Kremes As Ri adalah soto, ayam, dan kremes. Restoran La Moda Del Gelato hanya memiliki satu kata yang serupa yaitu gelato. Eliminasi dilakukan dengan cara menghilangkan kata-kata tersebut. *Item profile* yang telah melalui proses eliminasi ditunjukkan pada Tabel 5.

#### 4.3.4 Filtering

Tahap akhir *preprocessing* adalah *filtering*. Tahap *filtering* dilakukan dengan cara menghilangkan kata dengan panjang kurang dari atau sama dengan tiga huruf pada *Item profile*

hasil eliminasi. *Filtering* dilakukan karena kata-kata tersebut dianggap tidak penting dan dapat mengganggu proses selanjutnya yaitu perhitungan similaritas. Pengecualian dilakukan terhadap kata-kata yang termasuk ke dalam *wordlist*. Kata yang termasuk ke dalam *wordlist* ditunjukkan pada Tabel 6.

Tabel 4 *Item Profile* hasil tokenisasi

No	<i>Item Profile</i>		
	Ayam Geprek Keprabon	Soto Ayam Kremes As Ri	La Moda Del Gelato
1	ayam	soto	la
2	geprek	ayam	moda
3	keprabon	kremes	del
4	ayam	as	gelato
5	kremes	ri	chocolate
6	cappuccino	es	muffin
7	chocolate	buah	gelato
8	choco	susu	ice
9	cookies	keju	cream
10	paket	kremes	quiche
11	cup	paket	lorraine
12	ayam	b	red
13	geprek	soto	velvet
14	blenger	ayam	cake
15	paket	komplrit	manis
16	cup	manis	halal
17	ayam	asin	ac
18	geprek	halal	wifi
19	jumbo	delivery	khas
20	paket	toilet	indonesia
21	geprek	ac	
22	blenger	khas	
23	keju	indonesia	
24	mozarella	kartasura	
25	paket		
26	geprek		
27	jumbo		
28	taro		
29	smoothies		
30	pedas		
31	manis		
32	halal		
33	delivery		
34	toilet		
35	ac		
36	khas		
37	indonesia		
38	keprabon		

Tabel 5 *Item profile* hasil eliminasi

No	<i>Item Profile</i>		
	Ayam Geprek Keprabon	Soto Ayam Kremes As Ri	La Moda Del Gelato
1	ayam	soto	la
2	geprek	ayam	moda
3	keprabon	kremes	del
4	kremes	as	gelato
5	cappuccino	ri	chocolate
6	chocolate	es	muffin
7	choco	buah	ice
8	cookies	susu	cream
9	paket	keju	red
10	cup	paket	velvet

11	blenger	b	cake
12	keju	komplrit	quinche
13	mozarella	manis	lorraine
14	jumbo	asin	khas
15	taro	halal	indonesia
16	smoothies	delivery	
17	pedas	toilet	
18	manis	ac	
19	halal	tv	
20	delivery	khas	
21	toilet	indonesia	
22	ac	kartasura	
23	khas		
24	indonesia		

Tabel 6 *Wordlist*

<i>Wordlist</i>			
uns,	ums,	ac,	ice,
jus,	mie,	hik,	hot,
isi,	teh,	tv,	cup,
ati,	sup,	es,	iga,
air,	mix,	bun,	box,
red,	iso,	egg,	kfc,
bbq,	ijo,	kue,	sop,
kol,	ham,	fla,	pie,
gum,	ubi,	fry,	ebi

Tabel 7 *Item profile* hasil *preprocessing*

No	<i>Item Profile</i>		
	Ayam Geprek Keprabon	Soto Ayam Kremes As Ri	La Moda Del Gelato
1	ayam	soto	moda
2	geprek	ayam	gelato
3	keprabon	kremes	chocolate
4	kremes	es	muffin
5	cappuccino	buah	ice
6	chocolate	susu	cream
7	choco	keju	red
8	cookies	paket	velvet
9	paket	komplrit	cake
10	cup	manis	quinche
11	blenger	asin	lorraine
12	keju	halal	khas
13	mozarella	delivery	indonesia
14	jumbo	toilet	
15	taro	ac	
16	smoothies	tv	
17	pedas	khas	
18	manis	indonesia	
19	halal	kartasura	
20	delivery		
21	toilet		
22	ac		
23	khas		
24	indonesia		

Tidak ada kata yang difilter pada *item profile* Ayam Geprek Keprabon. Kata “cup” dan “ac” memang kurang dari atau sama dengan tiga huruf, namun kata tersebut termasuk ke dalam *wordlist*, sehingga tidak ikut difilter.

Tiga kata yang difilter pada *item profile* Soto Kremes As Ri yaitu “as”, “ri”, dan “b” karena kata tersebut kurang dari tiga huruf dan tidak termasuk ke dalam *wordlist*. Kata “la” dan “del” difilter dari *item profile* La Moda Del Gelato. *Item profile* yang telah melalui tahap *filtering* diasumsikan telah sesuai karena telah melalui seluruh tahapan *preprocessing*. *Item profile* hasil *preprocessing* ditunjukkan pada Tabel 7.

## 4.4 Implementasi Content-Based Filtering

Metode *content-based filtering* memiliki prinsip memberikan rekomendasi berdasarkan kemiripan *item profile*. *Item profile* yang sudah sesuai setelah melalui tahap *preprocessing* selanjutnya dibandingkan untuk dicari kemiripannya. Kemiripan dihitung dengan menggunakan fungsi *cosine similarity*.

### 4.2.1 Perhitungan Similaritas

Perhitungan yang dilakukan pertama kali adalah membandingkan *item profile* Ayam Geprek Keprabon dengan *item profile* Soto Ayam Kremes As Ri. Masing-masing *item profile* kedua restoran tersebut pada Tabel 7 dibandingkan dan ditemukan beberapa kemiripan seperti pada Tabel 8.

Tabel 8 Kemiripan *Item Profile*

No	<i>Item Profile</i>	
	Ayam Geprek Keprabon	Soyo Ayam Kremes As Ri
1	ayam	ayam
2	kremes	kremes
3	paket	paket
4	manis	manis
5	keju	keju
6	halal	halal
7	delivery	delivery
8	toilet	toilet
9	ac	ac
10	khas	khas
11	indonesia	indonesia

1. Total fitur konten dari *item profile* Ayam Geprek Keprabon,  $n(A)=24$ .
2. Total fitur konten dari *item profile* Soto Ayam Kremes As Ri,  $n(B)=19$ .
3. Total kemiripan fitur konten dari kedua *item profile* restoran tersebut adalah,  $n(A \cap B)=11$ .

Nilai similaritas dihitung menggunakan Persamaan (1).

$$\text{sim}(A, B) = \frac{11}{\sqrt{24 \cdot 19}}$$

$$\text{sim}(A, B) = 0,5151$$

Nilai similaritas untuk kedua restoran tersebut adalah 0,5151.

Perhitungan kedua adalah membandingkan *item profile* Ayam Geprek Keprabon dengan *item profile* La Moda Del Gelato. Kedua *item profile* pada Tabel 7 untuk restoran tersebut dibandingkan dan ditemukan beberapa keyword yang mirip seperti pada Tabel 9.

Tabel 9 Kemiripan *Item Profile*

No	<i>Item Profile</i>	
	Ayam Geprek Keprabon	La Moda Del Gelato
1	chocolate	chocolate
2	khas	khas
3	indonesia	indonesia

1. Total fitur konten dari *item profile* Ayam Geprek Keprabon,  $n(A)=24$ .

2. Total fitur konten dari *item profile* La Moda Del Gelato,  $n(B)=13$ .
3. Total kemiripan fitur konten dari kedua *item profile* restoran tersebut adalah,  $n(A \cap B)=3$ .  
Nilai similaritas dihitung dengan menggunakan Persamaan (1).

$$\text{sim}(A, B) = \frac{3}{\sqrt{24 \cdot 13}}$$

$$\text{sim}(A, B) = 0,1698$$

Tinggi rendahnya nilai similaritas dipengaruhi oleh kemiripan *item profile* antar restoran. Restoran Soto Ayam Kremes As Ri memiliki nilai similaritas tertinggi yakni 0,5151 karena *item profile* nya mirip dengan *item profile* restoran Ayam Geprek Keprabon, sedangkan restoran La Moda Del Gelato memiliki nilai similaritas yang sangat rendah yaitu sebesar 0,1698 karena hanya ditemukan 3 kemiripan fitur konten ketika *item profile* keduanya dibandingkan. Fitur konten yang terdapat pada restoran La Moda Del Gelato juga masih belum lengkap, sehingga menyebabkan hasil perhitungan similaritas sangat rendah.

### 4.2.2 Top-N Recommendation

Implementasi terus dilakukan untuk semua restoran hingga akhirnya didapat semua nilai similaritasnya. Ketika pengguna memilih restoran Ayam Geprek Keprabon, pengguna diberikan *top-N recommendation* urut dari nilai similaritas yang tertinggi. Lima rekomendasi yang diberikan kepada pengguna tersaji pada Tabel 10.

Tabel 10 *Top-5 Recommendation*

Nama Resto	Nilai Similaritas
Soto Ayam Kremes As Ri	0,5151
Warung Makan Busardjo	0,3969
Ayam Tentrem Sakabehe	0,3916
Ayam Zuprek	0,3849
Ayam Geprek Abang Ireng	0,3818

Tabel 10 menunjukkan bahwa restoran Soto Ayam Kremes As Ri masuk ke dalam *top-N recommendation* karena memiliki nilai similaritas tertinggi, sedangkan restoran La Moda Del Gelato tidak termasuk ke dalam hasil rekomendasi karena nilai similaritasnya sangat rendah.

## 4.5 Evaluasi

Hasil rekomendasi selanjutnya diuji presisi dan akurasi. Presisi dan akurasi diuji dengan menggunakan tiga *threshold* yang berbeda. Ketiga parameter *threshold* yang digunakan dalam pengujian ini adalah 0,33; 0,35; dan 0,38. Nilai parameter *threshold* tersebut ditentukan atas dasar empiris. Jika *threshold* terlalu rendah, maka restoran yang muncul sebagai rekomendasi terlalu banyak, begitu pula jika terlalu tinggi, maka restoran yang direkomendasikan terlalu sedikit.

Pengujian dilakukan dengan cara membandingkan kesesuaian hasil rekomendasi restoran Ayam Geprek Keprabon dengan restoran yang termasuk pada kategori Khas Indonesia. Kategori Khas Indonesia dijadikan sebagai data pembandingan karena merupakan salah satu kriteria fitur konten dari restoran Ayam Geprek Keprabon dan pada aplikasi MANGAN sendiri belum terdapat sistem

rekomendasi sebagai bahan untuk dibandingkan. Daftar restoran yang termasuk dalam kategori Khas Indonesia sejumlah 179 restoran tersaji pada Tabel 11.

Tabel 11 Restoran Khas Indonesia

No	Restoran Khas Indonesia
1	Soto Ayam Kremes As Ri
2	Ayam Kendil Solo
3	Ceker Bledex Mbah Mono
.	.
.	.
.	.
177	Sambel mbok Ti
178	Serabi Notosuman Ny. Lidia
179	Wedangan Pak Item

Tabel 1 digunakan sebagai acuan dalam menghitung presisi dan akurasi.

1. *True Positif* (TP) dihitung 1 jika restoran hasil rekomendasi termasuk restoran Khas Indonesia.
2. *False Positif* (FP) dihitung 1 jika restoran hasil rekomendasi tidak termasuk pada kategori Khas Indonesia.
3. Restoran yang termasuk dalam kategori Khas Indonesia tetapi tidak ditampilkan dalam rekomendasi dihitung sebagai *False Negative* (FN).
4. Restoran yang sama sekali tidak termasuk dalam hasil rekomendasi maupun kategori Khas Indonesia dimasukkan ke dalam *True Negative* (TN).

Rekomendasi yang dihasilkan untuk masing-masing parameter *threshold* yang berbeda ditunjukkan pada Tabel 12. Semakin besar nilai *threshold*, maka daftar restoran yang akan direkomendasikan akan semakin sedikit karena restoran dengan nilai similaritas dibawah *threshold* tidak ditampilkan.

Tabel 12 Hasil rekomendasi

No	Hasil Rekomendasi Restoran Ayam Geprek Keprabon		
	<i>Threshold</i> 0,33	<i>Threshold</i> 0,35	<i>Threshold</i> 0,38
1	Soto Ayam Kremes As Ri	Soto Ayam Kremes As Ri	Soto Ayam Kremes As Ri
2	Warung Makan Busardjo	Warung Makan Busardjo	Warung Makan Busardjo
3	Ayam Tentrem Sakabehe	Ayam Tentrem Sakabehe	Ayam Tentrem Sakabehe
4	Ayam Zuprek	Ayam Zuprek	Ayam Zuprek
5	Geprek Abang Ireng	Geprek Abang Ireng	Geprek Abang Ireng
6	Ayam Geprek Kumlot	Ayam Geprek Kumlot	
7	Waroeng TOP Inspirasi	Waroeng TOP Inspirasi	
8	Ayam-Ayam Resto	Ayam-Ayam Resto	
9	Sate Ayam Madura Cak Doni	Sate Ayam Madura Cak Doni	

10	Warung Makan Mas Elang		
11	Orlet's Solo		
12	Naluri Reborn		
13	Lesehan "ALDAN"		
14	Chicken Hut		

Pengujian pertama dilakukan dengan menggunakan *threshold* 0,33. Restoran dengan nilai similaritas dibawah 0,33 tidak ditampilkan sebagai rekomendasi, sehingga didapatkan 14 daftar rekomendasi restoran seperti yang tersaji pada Tabel 12. Hasil rekomendasi yang dihasilkan selanjutnya dibandingkan dengan Tabel 11.

Hasil perbandingan menunjukkan bahwa:

1. Restoran hasil rekomendasi yang termasuk ke dalam kategori Khas Indonesia ada 11.
2. Restoran hasil rekomendasi tidak termasuk pada kategori Khas Indonesia ada 3, yakni Sate Ayam Madura Cak Doni, Orlet's Solo, dan Chicken Hut.
3. Restoran yang termasuk dalam kategori Khas Indonesia tetapi tidak ditampilkan sebagai rekomendasi sebanyak 168.
4. Restoran yang sama sekali tidak termasuk dalam hasil rekomendasi maupun kategori Khas Indonesia ada 171.

Nilai hasil perbandingan selanjutnya dimasukkan ke dalam Tabel *confusion matrix* sehingga didapatkan nilai seperti pada Tabel 13.

Tabel 13 Confusion matrix *threshold* 0,33

		Hasil Sebenarnya	
		TRUE	FALSE
Prediksi	TRUE	11	3
	FALSE	168	171

Persamaan (2) dan Persamaan (3) digunakan untuk menghitung presisi dan akurasi:

$$presisi = \frac{11}{11 + 3} = 0,7857$$

$$akurasi = \frac{20 + 171}{20 + 171 + 3 + 159} = 0,5155$$

Presisi dan akurasi untuk pengujian pertama masing-masing bernilai sebesar 0,7857 dan 0,5155.

Cara yang sama dilakukan untuk pengujian kedua dan ketiga sehingga didapatkan nilai rata-rata presisi dan akurasi seperti pada Tabel 14.

Tabel 14 Hasil Pengujian

Pengujian Ke-	<i>Threshold</i>	Presisi	Akurasi
1	0,33	0,7857	0,5155
2	0,35	0,8888	0,5127
3	0,38	1	0,5070
Rata-rata		0,8915	0,5118

Rata-rata presisi dari ketiga pengujian yang telah dilakukan bernilai sebesar 0,8915. Ketika pengguna memilih restoran Ayam Geprek Keprabon yang termasuk ke dalam kategori Khas Indonesia, maka hasil yang diharapkan adalah sistem akan memberikan rekomendasi restoran-restoran yang termasuk dalam kategori Khas Indonesia juga. Ketiga pengujian yang telah dilakukan menunjukkan bahwa sistem berhasil memberikan rekomendasi restoran yang termasuk ke dalam kategori Khas Indonesia.

Tingginya nilai presisi, namun rendahnya nilai akurasi yang hanya sebesar 0,5118 menunjukkan bahwa terjadi kesalahan sistematis. Salah satu faktor yang mempengaruhi rendahnya nilai akurasi pada pengujian ini adalah jumlah data. Jumlah restoran yang terdaftar pada aplikasi MANGAN masih terlalu sedikit. Metode pengujian yang kurang tepat juga menjadi faktor lain dalam kesalahan sistematis sehingga menghasilkan bias yang cukup besar terhadap tingkat kedekatan antara nilai prediksi dengan nilai aktual

## 5. KESIMPULAN & SARAN

### 5.1 Kesimpulan

Penelitian yang telah dilakukan menunjukkan bahwa metode *content-based filtering* dapat diterapkan pada sistem rekomendasi untuk studi kasus aplikasi MANGAN. Hasil rekomendasi dipengaruhi oleh kemiripan *item profile* restoran yang dipilih oleh pengguna dengan *item profile* restoran lainnya. *Item profile*. Kelengkapan fitur konten juga ikut mempengaruhi hasil similaritas. Ketika restoran Ayam Geprek Keprabon dipilih, maka restoran Soto Ayam Kremes As Ri direkomendasikan kepada pengguna karena *item profile* nya mirip dengan *item profile* restoran Ayam Geprek Keprabon yakni dengan nilai similaritas sebesar 0,5151. Restoran La Moda Del Gelato tidak direkomendasikan karena nilai similaritasnya rendah yakni 0,1698.

Pengujian dilakukan dengan cara membandingkan hasil rekomendasi dengan data pembanding restoran pada kategori Khas Indonesia yang merupakan salah satu kriteria dari fitur konten pada *item profile* Ayam Geprek Keprabon. Hasil pengujian menunjukkan bahwa rekomendasi yang dibangkitkan memiliki rata-rata nilai presisi sebesar 0,8915 dan rata-rata nilai akurasi sebesar 0,5118. Tingginya nilai akurasi menunjukkan bahwa sistem tepat dalam memberikan jawaban terhadap informasi yang dibutuhkan oleh pengguna. Rendahnya nilai akurasi terjadi karena terdapat kesalahan sistematis karena faktor jumlah data yang diuji masih terlalu sedikit dan metode pengujian yang dilakukan masih kurang maksimal.

### 5.2 Saran

Perhitungan similaritas yang dilakukan masih per kata, sehingga rekomendasi yang dihasilkan masih kurang tepat. Penelitian selanjutnya dapat dikembangkan dengan cara menghitung nilai similaritas per kalimat atau bisa juga dengan menggunakan bobot agar hasil rekomendasi yang dihasilkan lebih relevan.

## DAFTAR PUSTAKA

- [1] S. Kangas, "Collaborative filtering and recommendation systems," *VTT Inf. Technol.*, vol. 1, pp. 1–34, 2001.
- [2] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," *Proc. 10th ...*, vol. 1, pp. 285–295, 2001.
- [3] F. Ricci, L. Rokach, and B. Shapira, "Introduction to Recommender Systems Handbook," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Boston, MA: Springer US, 2011, pp. 1–35.
- [4] M. Pazzani and D. Billsus, "Content-based recommendation systems," *Adapt. web*, vol. 10, no. 1, pp. 1–18, 2007.
- [5] H. Maharani and F. A. Gunawan, "Sistem Rekomendasi Mobil Berdasarkan Demographic dan Content-Based Filtering," *J. Telemat.*, vol. 9, no. 2, pp. 64–68, 2014.
- [6] A. A. Huda, "Pemanfaatan Karakter Rasa pada Makanan sebagai Dasar Penentuan Rekomendasi Restoran," *J. Inform. Sunan Kalijaga*, vol. 1, no. 2, pp. 58–66, 2016.
- [7] R. I. Victor, Adelia, M. Ayub, and A. Rehatta, "Implementasi Cosine Similarity dan Algoritma Smith-Waterman untuk Mendeteksi Kemiripan Teks," *J. Inform.*, vol. 10, no. 1, pp. 31–42, 2014.
- [8] S. A. Karunia, R. Saptono, and R. Anggrainingsih, "Online News Classification Using Naive Bayes Classifier with Mutual Information for Feature Selection," *J. Ilm. Teknol. dan Inf.*, vol. 6, no. 1, pp. 10–15, 2017.
- [9] R. Ranovan, A. Doewes, and R. Saptono, "Twitter Data Classification using Multinomial Naive Bayes for Tropical Diseases Mapping in Indonesia," *J. Telecommun. Electron. Comput. Eng.*, vol. 10, no. 2–4, pp. 155–159, 2018.
- [10] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, *Recommender systems: an introduction*, vol. 40. 2011.
- [11] L. Djumiroh and R. Saptono, "Penerapan Metode Collaborative Filtering Menggunakan Rating Implisit pada Sistem Rekomendasi Pemilihan Film di Rental VCD," *J. ITSMART*, vol. 1, no. 2, pp. 54–59, 2012.
- [12] E. Prasetyo, *Data Mining: Konsep Dan Aplikasi Menggunakan Matlab*, 1st Pub. Yogyakarta: ANDI Publisher, 2013.
- [13] B. Raharjo, "Nature, Environment, Remote Sensing, GIS, IT and Myself," 2011. [Online]. Available: <https://beniraharjo.wordpress.com/2011/03/17/presisi-dan-akurasi/>. [Accessed: 12-Apr-2018].