

Rekomendasi Penjual Bahan Makanan Dengan Metode Filtering Berbasis Konten dan Lokasi Pada Aplikasi Resep Masakan

Didik Irawan¹, Kusri², Muhammad Rudyanto Arief³

Magister Teknik Informatika, Universitas AMIKOM Yogyakarta^{1,2,3}
didikirawan@students.amikom.ac.id, kusri@amikom.ac.id, rudy@amikom.ac.id

Abstrak – Informasi resep masakan sangat membantu masyarakat dalam membuat menu masakan yang variatif setiap harinya. Berbagai aplikasi resep masakan dapat diunduh di internet dan pengguna dapat berbagi informasi mengenai inovasi dan variasi lain dari sebuah resep masakan. Namun satu hal yang menjadi keterbatasannya adalah tidak adanya informasi pendukung terkait dengan tempat untuk membeli bahan makanan yang ada pada resep tersebut. Setiap pengguna memiliki lokasi yang berbeda-beda, sehingga untuk mendapatkan bahan yang sama memerlukan pencarian yang tidak mudah. Berbagai pendekatan diteliti untuk dapat menjawab permasalahan tersebut, salah satunya dengan penggabungan metode *content based-filtering* dan *location based-filtering*. Pada penelitian ini, penulis menerapkan algoritma *text processing* dan TF-IDF untuk filtering konten dan untuk menerapkan algoritma klasifikasi *K-Nearest Neighbors* (KNN) dan formula *haversine* untuk mencari irisan antara bahan masakan dan jarak lokasi penjual terdekat yang menjual sebagian besar atau seluruh bahan pada resep masakan. Hasil dari penelitian ini menunjukkan bahwa penggabungan algoritma *text processing*, TF-IDF, KNN dan formula *haversine* dapat digunakan untuk menemukan penjual bahan makanan terdekat yang sesuai dengan resep masakan.

Kata kunci: *content based-filtering; location based-filtering; knn; tf-idf; haversine;*

Abstract - Recipe information is very helpful for the community in making a varied menu of dishes every day. Various recipe applications can be downloaded on the internet and users can share information about innovations and other variations of a recipe. However, one limitation is no supporting information regarding where to buy the ingredients in the recipe. Each user has a different location, so finding the same material requires a difficult search. Various approaches have been investigated to answer these problems, one of which is by combining content-based filtering and location-based filtering methods. In this study, the authors applied the text processing algorithm and TF-IDF for content filtering to apply for K-Nearest Neighbors (KNN) classification algorithm and the haversine formula to find the slices between cooking ingredients and the distance to the location of the closest seller who sells most or all of the ingredients in recipes. The results of this study indicate that the combination of the text processing algorithm, TF-IDF, KNN and the haversine formula can be used to find the closest food seller according to the recipe.

Keyword : *content based-filtering; location based-filtering; knn; tf-idf; haversine;*

1. Latar Belakang

Informasi resep masakan sangat membantu masyarakat dalam membuat menu masakan yang variatif setiap harinya. Tidak menutup kemungkinan bahwa menu masakan rumah akan monoton dan hanya memiliki sedikit variasi tanpa adanya informasi resep masakan. Resep masakan yang dulu dapat dibaca dari koran, majalah dan buku kini sudah beralih kepada informasi digital seiring perkembangan jaman. Berbagai aplikasi resep masakan dapat diunduh di internet dan pengguna dapat berbagi informasi mengenai inovasi dan variasi lain dari sebuah resep masakan.

Satu hal yang menjadi keterbatasan dari sebuah informasi resep masakan dari dulu

hingga saat ini adalah tidak adanya informasi pendukung terkait dengan tempat untuk membeli bahan makanan yang ada pada resep tersebut. Penjual di setiap daerah akan berbeda tergantung dengan lokasi pembaca berada. Pada aplikasi resep masakan, setiap pengguna memiliki lokasi yang berbeda-beda, sehingga untuk mendapatkan bahan yang sama memerlukan pencarian yang tidak mudah.

Berbagai pendekatan diteliti untuk dapat menjawab permasalahan tersebut, salah satunya dengan penggabungan metode *content based-filtering* dan *location based-filtering* untuk merekomendasi penjual terdekat. Sistem rekomendasi didefinisikan sebagai aplikasi pada website e-commerce

Rekomendasi Penjual Bahan Makanan Dengan Metode Filtering Berbasis Konten dan Lokasi
Pada Aplikasi Resep Masakan

untuk mengusulkan informasi dan menyediakan fasilitas yang diinginkan pengguna dalam membuat suatu keputusan [18]. Sistem tersebut diasumsikan sebagai representasi kebutuhan dan gambaran keinginan pengguna melalui pendekatan rekomendasi atau saran dari sebuah item atau informasi lain yang memiliki kemiripan dari karakteristik informasi.

Setiap metode yang digunakan pada *recommender system* memiliki kelebihan dan kekurangan, tidak terkecuali metode *content-based filtering* maupun *location-based filtering*. Untuk menutupi kelemahan pada metode-metode tersebut dapat dilakukan dengan menggunakan metode hybrid yaitu menggabungkan beberapa metode yang terdapat pada sistem rekomendasi untuk menghasilkan item rekomendasi yang relevan dengan keinginan pengguna.

Pada penelitian ini, penulis menerapkan algoritma text processing dan TF-IDF untuk filtering konten dan untuk menerapkan algoritma klasifikasi K-Nearest Neighbors (KNN) dan formula haversine. Apakah penggabungan metode tersebut dapat digunakan untuk mencari irisan antara bahan masakan dan jarak lokasi penjual terdekat yang menjual sebagian besar atau seluruh bahan pada resep masakan.

2. Kajian Pustaka

Penelitian ini dilakukan tidak terlepas dari penelitian yang pernah dilakukan sebelumnya dengan judul "Sistem Rekomendasi Pada E-Commerce Menggunakan K-Nearest Neighbor"[13], menyimpulkan metode Content Based memberikan hasil paling tinggi dibandingkan dengan Collaborative Filtering dan Hybrid pada $k=10$. Dari percobaan yang dilakukan pada ketiga metode apabila nilai k yang semakin tinggi, maka nilai recall yang dihasilkan juga semakin tinggi.

Penelitian lain dengan judul "Penerapan Metode Content Based Filtering Dalam Implementasi Sistem Rekomendasi Tanaman Pangan".[11] Tujuan dari penelitian ini adalah Memberi rekomendasi bagaimana cara memangkas rantai distribusi rantai pemasaran sehingga dapat meningkatkan kesejahteraan petani. Penelitian tersebut mampu membuktikan bahwa metode content-based filtering dapat digunakan untuk recommender system dengan hasil pengujian presisi rata-rata sebesar 78.40%. Namun dalam pengujian confusion matrix hanya perhitungan precision sehingga tidak diketahui besaran nilai accuracy dan recallnya.

Penelitian selanjutnya dengan judul "Content-based Recommender System for Movie Website"[8], dimana tesis tersebut bertujuan untuk menemukan cara yang baru untuk meningkatkan klasifikasi film dan meningkatkan rekomendasi berbasis konten film kepada penggunanya. Penelitian tersebut menggunakan TF-IDF-DC dan mengekstraksi fitur untuk mengevaluasi nilai akurasi. Dan selanjutnya menggunakan KNN dan berbagai metrik untuk mengevaluasi peningkatan pendekatan baru. Diilustrasikan bahwa pendekatan baru berkontribusi positif sesuai dengan evaluasi.

Kajian lain melakukan penelitian dengan judul "Pengujian Akurasi Sistem Rekomendasi Berbasis Content-Based Filtering"[12]. Penelitian tersebut meninjau pengaruh stopword dalam tingkat akurasi penentuan rekomendasi dokumen dengan klasifikasi dokumen yang digunakan. Penelitian tersebut menyimpulkan nilai sistem rekomendasi dengan proses stopword removal masih lebih unggul dibandingkan sistem rekomendasi tanpa proses stopword removal.

Selanjutnya, penelitian dengan judul "Pengaruh User Profiling Pada Rekomendasi Sistem Menggunakan K-Means dan KNN" [4] yang bertujuan untuk mengatasi masalah sparsity dan scalability dengan menggabungkan algoritma silhouette, k-means, K-Nearest Neighbour. Penelitian tersebut menyimpulkan bahwa Average Silhouette Method mampu mengoptimalkan fungsi K-Means dalam menentukan jumlah kluster dan Algoritma KNN memiliki performa cukup baik.

3. Metode Penelitian

a. Pengumpulan Data

Metode pengumpulan data dilakukan dengan cara observasi. Observasi dilakukan dengan pengamatan langsung terhadap objek yang akan diteliti terkait permasalahan yang akan dibahas. Objek yang akan diteliti yaitu data set resep masakan dan data toko atau warung penjual bahan makanan beserta barang jualan.

b. Preprocessing

Data yang telah dikumpulkan dalam bentuk dataset telah memiliki karakteristik sendiri. Data set yang ada berupa text dan harus melalui pengolahan text processing. Bahan-bahan yang ditulis dalam resep umumnya adalah kata baku sehingga beberapa text processing yang dilakukan adalah :

• **Case Folding**

Case Folding yang merupakan proses untuk merubah setiap kata menjadi sama, misal huruf kecil dengan menggunakan fungsi lowercase. Contoh proses Case Folding dapat dilihat pada **Tabel 1**.

Tabel 1. Case Folding

Input	Bahan-bahan yang dibutuhkan adalah Serai, Lengkuas, Jahe, Garam.
Case Folding	bahan-bahan yang dibutuhkan adalah serai, lengkuas, jahe, garam.

• **Tokenizing**

Tokenizing merupakan proses untuk memenggal setiap kata dalam kalimat atau pemotongan string input berdasarkan tiap kata yang menyusunnya. Tokenisasi secara garis besar memecah sekumpulan karakter dalam suatu teks ke dalam satuan kata, bagaimana membedakan karakter-karakter tertentu yang dapat diperlakukan sebagai pemisah kata atau bukan. Whitespace, enter, tab dianggap sebagai pemisah kata, dan karakter seperti titik (.), koma (,), semicolon (;), petik (') dan karakter lainnya memiliki peran cukup penting dalam pemisahan kata. Contoh Tokenizing dapat dilihat pada **Tabel 2**.

Tabel 2. Tokenizing

Hasil Case Folding	bahan-bahan yang dibutuhkan adalah serai, lengkuas, jahe, garam.
Tokenizing	bahan bahan yang dibutuhkan adalah serai lengkuas jahe garam

• **Filtering**

Tahap Filtering adalah tahap mengambil kata-kata penting dari hasil token. Terdapat dua cara dalam filtering, yaitu menggunakan algoritma stoplist (membuang kata kurang penting) seperti “yang”, “di”, “ke” dan “dari” atau wordlist (menyimpan kata penting). Dalam penelitian ini digunakan algoritma wordlist yaitu menyimpan kata-kata yang penting. Setiap data yang filter akan di cocokkan dengan kamus bahan makanan pada database untuk selanjutnya digunakan sebagai pembobotan. Contoh proses Filtering yang terjadi pada sistem warung online ditunjukkan oleh **Tabel 3**.

Tabel 3. Filtering

Tokenizing	bahan bahan yang dibutuhkan
------------	-----------------------------------

Rekomendasi Penjual Bahan Makanan Dengan Metode Filtering Berbasis Konten dan Lokasi Pada Aplikasi Resep Masakan

	adalah serai lengkuas jahe garam
Dictionary	cabai, gula, serai, bawang merah, garam, bawang putih, serai, lengkuas, jahe, ...dst
Hasil Filtering	serai lengkuas jahe garam

c. Pembobotan

Metode TF-IDF merupakan metode untuk menghitung bobot setiap kata yang paling umum digunakan pada information retrieval. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat. Metode ini akan menghitung nilai Term Frequency (TF) dan Inverse Document Frequency (IDF) pada setiap token (kata) di setiap dokumen dalam korpus [17]. Metode ini akan menghitung bobot setiap token t di dokumen d dengan **Persamaan 1**:

$$W_{dt} = tf_{dt} * IDF_t \tag{1}$$

Dimana :

d adalah dokumen ke-d

t adalah kata ke-t dari kata kunci

W adalah bobot dokumen ke-d terhadap kata ke-t

tf adalah banyaknya kata yang dicari pada sebuah dokumen

Setelah bobot (W) masing-masing dokumen diketahui, maka dilakukan proses pengurutan dimana semakin besar nilai W, semakin besar tingkat similaritas dokumen tersebut terhadap kata kunci, demikian sebaliknya. Algoritma Term Frequency-Invers Dokument Frequeency TF-IDF akan membangun profil berdasarkan bobot vector pada komponen pembentuknya. Berdasarkan preferensi pengguna maka sistem akan memilih warung berdasarkan analisis kemiripan dari informasi yang diakses pengguna untuk mencapai tujuan yang telah ditetapkan. Nilai IDF dapat dihitung berdasarkan **Persamaan 2**:

$$(idf_t)^n = \log\left(\frac{n}{idf_t}\right) \tag{2}$$

Dimana :

n = jumlah semua produk

df = n jumlah produk yang memiliki term i.

d. Klasifikasi

• **K-Nearest Neighbors**

Algoritma K-Nearest Neighbor merupakan salah satu metode yang digunakan dalam pengklasifikasian[17]. Prinsip kerja K-Nearest Neighbor (KNN) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (neighbor) terdekatnya dalam data pelatihan. Pada penelitian ini, digunakan rumus Euclidean Distance dimana seperti pada Persamaan 3 berikut :

$$d_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (3)$$

Dimana :

d_{ij} adalah jarak antara i dan j

n adalah banyaknya variabel yang di amati

x_{ik} adalah nilai obyek i pada variabel ke- k

x_{jk} adalah nilai obyek j pada variabel ke- k

• **Formula Haversine**

Haversine formula adalah metode perhitungan jarak antara dua titik di bumi berdasarkan panjang garis lurus antara dua titik tanpa mengabaikan kelengkungan bumi[10]. Berikut adalah rumus haversine pada **Persamaan 4**:

$$a = \sin^2\left(\frac{\Delta lat}{2}\right) + \cos(lat_1) \cdot \cos(lat_2) \cdot \sin^2\left(\frac{\Delta long}{2}\right) \quad (4)$$

$$d = 2r \cdot \arcsin(\sqrt{a})$$

Keterangan :

d = Jarak

r = Jari-jari bumi

Δlat = besaran perubahan latitude

$\Delta long$ = besaran perubahan longitude

4. Implementasi Sistem dan Hasil

a. Preprocessing

Sebuah resep masakan memiliki daftar makanan yang diproses menggunakan metode text processing pada **Tabel 4** sebagai berikut :

Tabel 4. Preprocessing

Text	1 butir Telur Utuh 100 gram Gula Pasir 140 ml Air 120 gram Tepung Beras 60 gram Tepung Maizena 30 gram Tepung Terigu 25 gram Susu Bubuk 1/2 sdt Vanilla Bubuk 1/2 sdt Garam
Case Folding	1 butir telur utuh 100 gram gula pasir 140 ml air 120 gram tepung beras

	60 gram tepung maizena 30 gram tepung terigu 25 gram susu bubuk 1/2 sdt vanilla bubuk 1/2 sdt garam
Tokenizing	1 butir telur utuh 100 gram gula pasir 140 ml air 120 gram tepung beras 60 gram tepung maizena 30 gram tepung terigu 25 gram susu bubuk 1/2 sdt vanilla bubuk 1/2 sdt garam
Dictionary	bawang merah, bawang putih, cabai merah, telur, gula pasir, tepung beras, lada, merica, tepung maizena, tepung terigu, susu bubuk, vanili bubuk, garam
Filtering	telur gula pasir tepung beras tepung maizena tepung terigu susu bubuk vanilla bubuk garam

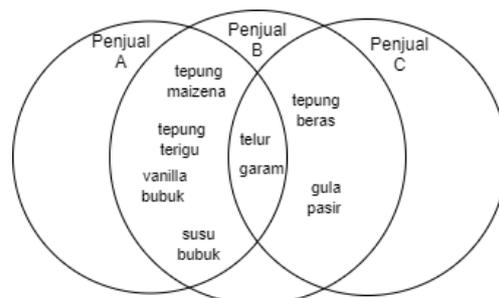
b. Pembobotan dan Klasifikasi

Pada tahap ini, hasil filtering akan diproses bersama data set penjual bahan. Didapatkan hasil pembobotan, perhitungan KNN dan perhitungan jarak menggunakan haversine sebagai berikut :

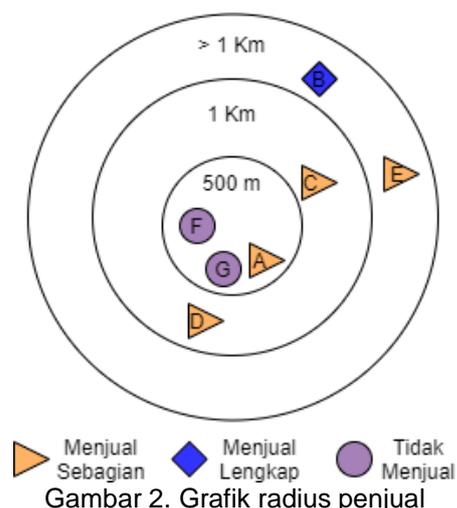
Tabel 5. Perhitungan

Penjual	TF-IDF	Modulus KNN K=10,20,50	Haversine
A	6.813	2	126 meter
B	-24.08	3	1,3 Km
C	9.166	3	617 meter
D	3.23	3	840 meter
E	8.33	2	1,5 Km
F	0	0	289 meter
G	0	0	150 meter

Dapat dilihat dari **Tabel 5** bahwa penjual B dengan jarak 1,3 Kilometer memiliki nilai TF-IDF paling kecil walau memiliki nilai modulus KNN yang sama dengan penjual C dan D. Visualisasi hubungan antara bahan makanan yang ada pada resep masakan dengan penjual dapat dilihat pada **gambar 1**.



Gambar 1. Contoh Irisan bahan antar penjual Visualisasi jarak penjual terhadap pengguna dapat dilihat pada **Gambar 2**.



Gambar 2. Grafik radius penjual

5. Kesimpulan

Perhitungan yang dilakukan pada 50 dataset menunjukkan bahwa Penjual B memiliki seluruh item yang ada pada resep masakan dengan jarak 1,3 Kilometer. Dengan algoritma TF-IDF dengan term bahan makanan, penjual B memiliki nilai -21.08. Penjual B juga muncul pada perhitungan algoritma KNN dengan nilai K 10, 20 dan 50. Dengan Semakin kecil nilai TF-IDF dan semakin besar nilai modulus KNN maka penjual tersebut memiliki sebagian besar menu yang ada pada resep masakan. Dari penelitian ini diperoleh kesimpulan bahwa penggabungan metode *content-based filtering* dengan *location based-filtering* dapat digunakan untuk melakukan rekomendasi penjual bahan makanan terdekat dari informasi resep masakan berupa text. Pendekatan yang digunakan adalah pendekatan klasifikasi menggunakan algoritma *text processing*, TF-IDF, KNN dan *haversine formula*.

Kekurangan dari penelitian ini adalah data yang masih sedikit, sehingga rekomendasi masih belum relevan untuk sebagian menu masakan. Untuk penelitian selanjutnya diharapkan menggunakan data yang lebih banyak dan dapat melakukan perhitungan nilai akurasi dari penggabungan metode *content-based filtering* dengan *location based-filtering (hybrid)*.

6. Pustaka

- [1] Achakulvisut, T., D. E. Acuna, T. Ruangrong, And K. Kording. 2016. "Science Concierge: A fast content based recommendation system for scientific," PLoS One. doi: doi.org/10.1371/journal.pone.0158423.
- [2] Bobadilla, J., F. Ortega, A. Hernando, And A. Gutiérrez. 2013. "Recommender Systems Survey. KnowledgeBased Systems," vol. 46, pp. 109–132.
- [3] Deshpande, M., A. V. Muddebihalkar, A. B. Jadhav, And S. Kokate. 2016. "Hybrid Content-Based Filtering Recommendation Algorithm on Hadoop," Int. J. Adv. Res. Comput. Eng. Technol., vol. 5.
- [4] Hartatik. Rosyid. 2020. Pengaruh User Profiling Pada Rekomendasi Sistem Menggunakan K-Means dan KNN. Jurnal Of Information System Management. Vol.2, No.1.
- [5] Kalita, J., V. E. Balas, S. Borah, And R. PRADHAN. 2019. "Recent Developments in Machine Learning and Data Analytics," Springer Singapore, vol. 740, 2019, doi: 10.1007/978-981-13-1280-9.
- [6] Keliwulan, Louis. 2019. "Perancangan dan Implementasi Website Sistem Rekomendasi Pemilihan Sepatu Menggunakan Algoritma Filtering dan Algoritma SAW". Universitas Teknologi Yogyakarta.
- [7] Kusriani., E T Luthfi. 2009. "Algoritma Data Mining". Yogyakarta: Penerbit Andi.
- [8] Ma, Ke. 2016. "Content-based Recommender System for Movie Website". KTH Royal Institute of Technology. Swedia.
- [9] Melita, Ria dkk. 2018. "Penerapan Metode Term Frequency Inverse Document Frequency (TF-IDF) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Syarah Umdatil Ahkam)". Jurnal Teknik Informatika Vol 11 No. 2. Universitas Islam Negeri Syarif Hidayatullah Jakarta.
- [10] Miftahudin, Yusuf dkk. 2020. Perbandingan Metode Perhitungan Jarak Euclidean, Haversine, Dan Manhattan Dalam Penentuan Posisi Karyawan (Studi Kasus : Institut Teknologi Nasional Bandung). Jurnal Tekno Insentif., vol. 14, no. 2, pp. 69-77, doi: 10.36787/jti.v14i2.270.
- [11] Nastiti, P. 2019. "Penerapan Metode Content Based Filtering Dalam Implementasi Sistem Rekomendasi Tanaman Pangan," Teknika, vol. 8, no. 1, pp. 1–10, 2019, doi: 10.34148/teknika.v8i1.139.

- [12] Parwita, Wgs. 2019. "Penguujian Akurasi Sistem Rekomendasi Berbasis Content-Based Filtering". *Jurnal Ilmiah Ilmu Komputer*, vol. 14, pp. 1, doi : 10.30872/jim.v14i1.1272
- [13] Prasetya, Candra S. D. 2017." Sistem Rekomendasi Pada E-Commerce Menggunakan K-Nearest Neighbor". *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*., Vol., No.3, hlm. 194-200.
- [14] Pratama, D. And S. Hansun. 2017. "Aplikasi Rekomendasi Tempat Makan Menggunakan Algoritma Slope One pada Platform Android," *IJCCS (Indonesian J. Comput. Cybern. Syst)*., vol. 11, no. 1, p. 11, 2017, doi: 10.22146/ijccs.15558.
- [15] Putu, N., R. Gangga, O. Sudana, And I. M. Sukarsa. 2017. "Implementasi Diagram Tree pada Rancang Bangun Sistem Informasi Bebayuhan Oton Berbasis Web". *Lontar Komput.*, vol. 8, no. 3, pp. 178–187.
- [16] Quercia, D., et al. 2010 "Recommending Social Events from Mobile Phone Location Data," 2010 IEEE International Conference on Data Mining, Sydney, NSW, 2010, pp. 971-976, doi: 10.1109/ICDM.2010.152.
- [17] Robertson, Stephen. 2005. Understanding Inverse Document Frequency: On theoretical arguments for IDF, England : *Journal of Documentation*, Vol. 60, pp. 502–520
- [18] Ricci, F. 2002. Travel recommender system. *IEEE Intelligent Systems* 17(6): 55-57. 2002
- [19] Ricii, F., et al., 2011. *Recommender System Handbook*, New York:Springer.
- [20] Setiawan, E., Soegiarto. 2017. "Aplikasi Rekomendasi Pemilihan Menu Makanan Sehat Untuk Anak Penderita Obesitas Menggunaka Metode Fuzzy Tahani". *Progresif: Jurnal Ilmiah Komputer*, Vol. 13, Hal.1.
- [21] Sutarman. 2012. "Pengantar Teknologi Informasi". Jakarta: Bumi Aksara.
- [22] Yuan, Q., Cong, G. & Lin, C.Y. 2014, August. COM: a generative model for group recommendation. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 163-172. ACM.