

K-MODES CLUSTERING UNTUK MENGETAHUI JENIS MASAKAN DAERAH YANG POPULER PADA WEBSITE RESEP ONLINE (STUDI KASUS: MASAKAN BANJAR DI COOKPAD.COM)

Fatma Indriani¹, Irwan Budiman²

^{1,2}Program Studi Ilmu Komputer,
Fakultas MIPA, Universitas Lambung Mangkurat
Email: ¹f.indriani@unlam.ac.id, ²irwan.budiman@unlam.ac.id

(Naskah masuk: 28 November 2017, diterima untuk diterbitkan: 30 Desember 2017)

Abstrak

Pada makalah ini dipaparkan clustering pada data resep masakan daerah Banjar untuk mengetahui jenis makanan yang paling banyak di-post secara online oleh pengguna website recipe sharing. Pertama-tama data resep sebanyak 355 dikumpulkan dari suatu website resep, untuk selanjutnya dilakukan ekstraksi data bahan dan pembersihan. Metode clustering yang dipilih adalah k-modes karena cocok digunakan pada data kategorikal. Berdasar metode Elbow, jumlah cluster yang ideal adalah k=4 dan k=8. Jumlah cluster k=4 menghasilkan kelompok yang lebih umum, sedangkan k=8 menghasilkan kelompok yang lebih spesifik. Adapun kelompok yang berhasil diidentifikasi untuk k=4 adalah sayur asam, soto banjar, masakan gurih lain-lain, kue dan bubur manis. Sedangkan kelompok dengan jumlah cluster k=8 adalah sayur asam, soto banjar, kue basah, masakan gurih lain-lain, masak habang, bubur manis, kuah ketupat, dan masakan gurih asam. Evaluasi nilai purity menunjukkan nilai masing-masing 0,825 untuk k=4 dan 0,831 untuk k=8.

Kata kunci: data mining, clustering, k-modes, resep masakan, bahan

Abstract

In this paper, we cluster user-submitted recipes of Banjar regional cuisine to find out which type of cuisine are popular according to its ingredients. 355 recipes are collected from a recipe sharing website, then the ingredients extracted and cleaned. The clustering method chosen is k-modes because it is suitable for categorical data. Based on the Elbow method, the ideal number of clusters is k = 4 and k = 8. The number of clusters k = 4 produces more general cuisines group, whereas k = 8 produces more specific groups. The groups identified for k = 4 are (1) "sayur asam" (sour soup), (2) "soto banjar" (Banjar chicken soup), (3) savory dishes, and (4) sweet dishes. While the group with the number of clusters k = 8 consists of (1) "sayur asam" (sour soup) (2) "soto banjar", (3) Banjar sweet puddings, (4) various savory dishes, (5) "masak habang" (Banjar sweet chili dishes), (6) sweet porridge, (7) "kuah ketupat" (spicy coconut soup) and (8) various savory sour dishes. The purity of clusters are shown to be 0.825 for k=4 and 0.831 for k=8.

Keywords: clustering, k-modes, data mining, recipe, ingredient

1. PENDAHULUAN

Indonesia memiliki tradisi masakan yang kaya di dunia (SBS Food, 2016). Ada lebih dari 6000 pulau di Indonesia, dan hampir tiap daerah memiliki makanan khas. Para pengguna internet banyak saling berbagi resep di website khusus resep masakan. Data resep tersebut tersedia secara bebas bisa dimanfaatkan untuk menggali informasi mengenai berbagai masakan yang populer di masyarakat, tak terkecuali masakan daerah. Informasi jenis masakan yang populer bisa digunakan untuk memantau tren pengguna terkait makanan.

Permasalahannya, karena resep di internet merupakan *user-generated content*, masakan yang sama bisa di-posting dengan berbagai versi, baik itu judul yang berbeda, bahan yang berbeda, maupun langkah-langkah memasak yang berbeda. Misal terdapat perbedaan antara resep soto pengguna A

dengan soto versi pengguna B. Dengan demikian, mencari jenis masakan yang banyak dibagikan tidak bisa dengan menghitung secara sederhana, karena perlu diidentifikasi masakan-masakan yang mirip tersebut ke dalam kelompok masing-masing terlebih dahulu.

Salah satu teknik penggalian pola adalah analisis *cluster*. Clustering bermaksud mengelompokkan titik data yang mirip. Pada makalah ini, akan disampaikan cara penggalian kelompok jenis masakan daerah menggunakan teknik *clustering*. Metode clustering yang dipilih adalah *k-modes* karena cocok untuk data kategorikal seperti pada bahan resep masakan. Fitur yang digunakan adalah bahan makanan, dengan asumsi bahwa jenis masakan yang sama akan memiliki bahan yang mirip pula. Studi kasus yang dipilih adalah masakan daerah Banjar (Kalimantan Selatan) yang tersedia pakar di daerah asal penulis untuk menginterpretasi dan

menvalidasi hasil *cluster*. Namun metodologi yang digunakan bisa diterapkan pada masakan daerah manapun.

Selain mendapatkan kelompok-kelompok masakan yang sejenis, hasil cluster bisa digunakan untuk mengidentifikasi kombinasi bahan yang sering digunakan berdasar sentroid *cluster*. Ini akan menunjukkan ciri khas bahan masakan suatu daerah tersebut. Kita bisa lebih memahami budaya kulinernya, bisa membuat masakan tersebut, bahkan bisa mengkreasikan masakan baru yang sesuai dengan ciri khas daerah tersebut.

Pada bagian 2 disampaikan penelitian terdahulu terkait dataset resep. Selanjutnya di bagian 3 dipaparkan metodologi yang dilakukan untuk menerapkan *k-modes* pada data resep. Di bagian 4 disampaikan hasil eksperimen, meliputi clustering dengan berbagai nilai *k*, serta penjelasan cluster dengan *k* terbaik. Dari hasil cluster akan didapat jenis resep yang sering dibagikan.

2. PENELITIAN TERDAHULU

Dalam ranah data mining, penelitian dalam pengolahan data resep pernah dilakukan melibatkan teknik *supervised* maupun *unsupervised learning*. Penerapan teknik *supervised learning* dilakukan oleh Su, dkk (2014), yaitu pembuatan model klasifikasi asal daerah suatu resep masakan menggunakan SVM dan SVD. Hasilnya diaplikasikan pada suatu sistem yang bisa memberi label asal daerah suatu masakan secara otomatis berdasarkan bahan-bahannya. Dari segi bahasa, Su menggunakan resep berbahasa Inggris, dan menggali karakteristik masakan khas daerah secara luas (negara), misal masakan Cina, Italia, Thailand, dll.

Teknik *supervised learning* juga diterapkan oleh Rufaidah (2016) yang membuat model klasifikasi resep MPASI (Makanan Pendamping ASI) ke dalam target kelompok umur, dengan metode ID3. Hasilnya berupa 40 aturan klasifikasi dengan akurasi rata-rata sebesar 73%.

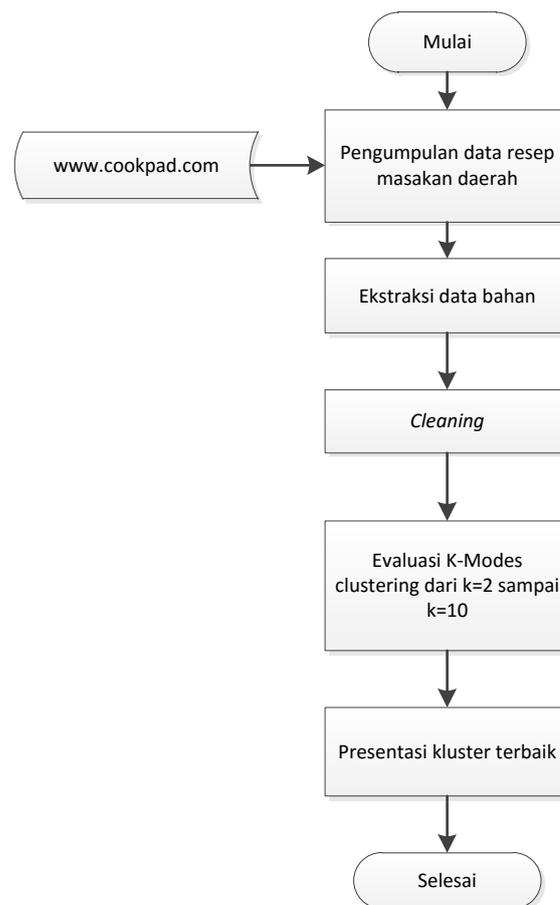
Sedangkan penelitian yang menerapkan teknik *unsupervised learning* yaitu Jain dkk (2015). Pada penelitian tersebut, dilakukan pencarian pasangan bahan yang sering ditemukan pada masakan khas India untuk dianalisis berdasar *flavor-compound* yang terkandung pada bahan-bahannya. Jain, dkk menemukan bahwa pada masakan India, sering terjadi pasangan bahan yang bertolak belakang dari segi rasa dan aroma. Konsep jaringan flavor dan prinsip pasangan bahan makanan diperkenalkan oleh Ahn, dkk (2011).

Nadamoto dkk (2016) menerapkan teknik *clustering* pada data resep dengan tujuan mengelompokkan hasil *search query* pada website resep. Teknik yang diusulkan terdiri dari dua tahap *clustering*, tahap pertama *clustering* berdasar bahan utama, tahap kedua *clustering* berdasar bumbu utama. Teknik ini berhasil mengelompokkan resep hasil pencarian ke dalam kelompok-kelompok yang mirip

untuk memperbaiki antarmuka pengguna pada hasil pencarian.

3. METODOLOGI

Langkah yang dilakukan meliputi pengumpulan data resep masakan daerah dari *website sharing* resep cookpad.com, ekstraksi bahan dan *cleaning*, menentukan nilai *k* pada *k-modes* dengan prinsip *Elbow method*, serta presentasi hasil *clustering* seperti yang dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alur Proses Clustering

3.1. Pengumpulan data

Pada tahap ini, dilakukan *crawling* website resep online Cookpad.com. *Query string* yang digunakan yaitu “masakan Banjar”, dan jumlah resep yang terkumpulkan adalah sebanyak 355.

3.2. Prapemrosesan Data

Pada tahap ini, data resep yang telah dikumpulkan akan melalui beberapa tahap prapemrosesan, yaitu:

- Ekstraksi bahan (*ingredient*). Data resep terdiri dari beberapa komponen, yaitu judul, link, daftar bahan beserta kuantitas, daftar langkah memasak, serta image gambar. Hanya daftar bahan yang

digunakan pada penelitian ini. Komponen lainnya diabaikan. Berikut ini contoh data yang didapat (Tabel 1).

Tabel 1. Contoh Hasil Ekstraksi Bahan dari Dataset Resep

<p>Sayur/Gangan/Jangan Asem khas Banjar https://cookpad.com/id/resep/2144448-sayurganganjangan-asem-khas-banjar patin ukuran kecil;mentimun;kol kecil;tomat;lombok besar;bawang merah besar;kemiri;kelingking laos;kelingking kunyit;serai;cabe rawit (sesuai selera);garam;gula pasir;saos tiram;air (panci uk. sedang)</p>
<p>Soto Banjar https://cookpad.com/id/resep/2135070-soto-banjar ayam (potong2);air;telur bebek rebus (ambil kuningnya aja);kentang uk kecil (rebus);Minyak untuk menumis;Garam+gula+kaldu (sesuai selera);Bumbu kering ;bunga lawang/pekak;kapulaga (tekan dengan jari sampai terbuka);cengkeh;kayu manis;Bumbu halus ;bawang merah;bawang putih;merica butiran;serutan buah pala;jahe;Pelengkap (sesuai selera);Ketupat/lontong;Soun/bihun rebus;Telur rebus;Sambal;Jeruk nipis;Seledri +prei/daun bawang;Bawang goreng;Kecap manis</p>

2.	cabai merah	cabe merah
	cabai	
	cabe	
3.	kaminting	kemiri
	kemiri	
4.	sereh	serai
	serai	
	sarai	
5.	lengkuas	lengkuas
	laos	
6.	telur	telur
	telor	
	hintalu	

Tabel 3. Contoh Hasil *Cleaning*

<p>Sayur/Gangan/Jangan Asem khas Banjar https://cookpad.com/id/resep/2144448-sayurganganjangan-asem-khas-banjar patin;timun;kol;tomat;cabe;bawang merah;kemiri;lengkuas;kunyit;serai;cabe rawit;cabe;garam;gula;saus tiram;panci</p>
<p>Soto Banjar https://cookpad.com/id/resep/2135070-soto-banjar kecap manis;bawang goreng;seledri;bawang prei;jeruk nipis;sambal;telur;suun bihun;ketupat lontong;jahe;pala;merica;bawang putih;bawang merah;kayu manis;cengkeh;kapulaga;bunga lawang;garam;gula;minyak;kentang;t elur bebek;ayam</p>

- b. *Cleaning* data meliputi (1) penghapusan karakter non huruf, (2) penghapusan *stopword* berbahasa Indonesia yang dibangun secara manual untuk data resep, serta (3) penyeragaman ejaan dan istilah.

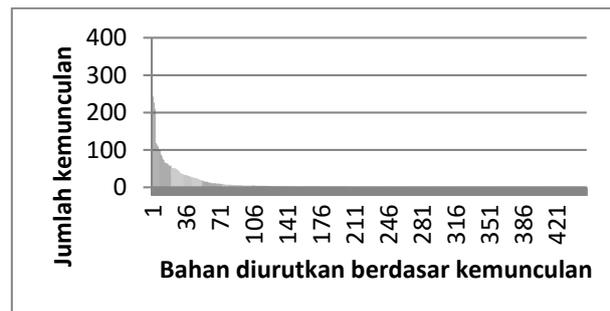
Stopword yang dibuat adalah kata-kata yang tidak ada hubungan dengan nama bahan namun sering muncul pada keterangan bahan, seperti “ada”, “ingin”, “jika”, “sedikit”, “tambahan”. Selain itu kata-kata keterangan yang sering muncul juga dimasukkan ke dalam stoplist, seperti “digeprek”, “diiris”, “dihaluskan”. Sedangkan penyeragaman ejaan dilakukan dengan sistem kamus yang dibuat secara manual (Tabel 2).

Jumlah fitur bahan akhir yaitu sebanyak 449 bahan unik. Rata-rata tiap resep mengandung 12,23 macam bahan. Contoh hasil pada tahap ini dapat dilihat di Tabel 3.

Tabel 2. Contoh beberapa transformasi yang dilakukan dengan sistem kamus

	Kata	Penyeragaman
1.	bwg merah	bawang merah
	bawang merah	
	bw merah	
	bamer	

- c. Vektorisasi data menghasilkan matriks biner berukuran 355 x 449 dengan fitur bahan. Bahan yang muncul pada tiap resep akan bernilai 1, sedangkan bahan yang tidak muncul bernilai 0. Dari 449 bahan yang ada, masing-masing memiliki kemunculan yang bervariasi, dengan rata-rata kemunculan bahan 9,65 kali. Ada 244 bahan yang hanya muncul 1 kali, dan bahan dengan kemunculan terbanyak yaitu garam 287 kali. Berikut ini adalah grafik frekuensi kemunculan tiap bahan pada dataset resep, diurutkan dari yang paling sering muncul (Gambar 1). Tabel 4 menunjukkan daftar 20 bahan teratas.



Gambar 2. Frekuensi Kemunculan Bahan

Tabel 4. 20 Bahan Dengan Kemunculan Terbanyak

No	Nama Bahan	Jumlah kemunculan
1	garam	287
2	bawang merah	243
3	bawang putih	227
4	gula	210
5	santan	118
6	jahe	114
7	telur	109
8	gula merah	102
9	cabe merah	99
10	kayu manis	88
11	ayam	84
12	kemiri	76
13	kunyit	73
14	minyak goreng	67
15	cengkeh	65
16	serai	65
17	tomat	63
18	seledri	61
19	merica	58
20	teras	58

3.3. K-Modes Clustering

K-modes clustering pertama kali diperkenalkan oleh Huang (1998) sebagai suatu metode *clustering* yang dikembangkan dari metode *k-means*. Oleh karena itu *k-modes* efisien seperti *k-means* namun digunakan pada data yang bersifat kategorikal. Modifikasi yang dilakukan terhadap metode *k-means* yaitu:

1. Jarak antara dua titik data X dan Y adalah jumlah fitur pada X dan Y yang nilainya berbeda (*simple dissimilarity measure*), secara formal dirumuskan seperti berikut ini:

$$d_1(X, Y) = \sum_{j=1}^m \delta(x_j, y_j) \quad (1)$$

di mana

$$\delta(x_j, y_j) = \begin{cases} 0 & (x_j = y_j) \\ 1 & (x_j \neq y_j) \end{cases} \quad (2)$$

dengan

x_j dan y_j adalah nilai fitur ke- j dari data X dan Y , serta m adalah jumlah fitur.

2. Ubah rata-rata (*means*) menjadi modus (*modes*)
3. Menggunakan frekuensi untuk mencari modus

Tabel 5. Contoh mencari sentroid dari 3 buah data menggunakan modus

ID	bawang merah	ayam	santan	jahe
1	1	1	1	0
2	1	0	0	1
3	1	1	1	0
Sentroid	1	1	1	0

Sebagai contoh dengan tiga data dummy berikut ini (Tabel 5), titik sentroid dibangun dari modus tiap

fitur. Modus adalah nilai data yang paling banyak muncul. Pembentukan sentroid adalah dengan mencari modus tiap fitur.

Berikut ini langkah-langkah *k-modes* clustering berdasar (Huang, 2008):

1. Pilih modus awal sejumlah k
2. Alokasikan objek data pada *cluster* terdekat berdasar *simple dissimilarity measure*. Update tiap modus *cluster* setelah tiap alokasi.
3. Setelah semua objek data telah dialokasikan ke suatu *cluster*, periksa kembali nilai *dissimilarity* tiap objek terhadap modus. Jika suatu objek data ternyata modus terdekat berada pada *cluster* lain, pindahkan objek ke *cluster* yang sesuai dan update modus kedua *cluster*.
4. Ulangi langkah 3 sampai tidak ada objek data yang berubah *cluster*.

Untuk menentukan jumlah *cluster* optimal digunakan *Elbow method* Goutte, dkk (1999) namun dimodifikasi menggunakan *within cluster difference*. Dari hasil plotting *within cluster difference* pada berbagai nilai, prinsip *Elbow method* mengambil nilai k pada titik ketika nilainya tidak mengalami penurunan signifikan dengan penambahan nilai k .

$$y = \sum_{i=1}^k \sum_{j=1}^m d_1(x_j, x_c) \quad (3)$$

dengan

y = jumlah *within cluster difference*

k = jumlah *cluster*

m = jumlah anggota pada tiap *cluster*

c = sentroid dari *cluster*

d = *simple dissimilarity measure* sesuai rumus (1)

x = titik data

Selanjutnya, evaluasi *cluster* dilakukan secara eksternal menggunakan metrik *purity* (Tan, dkk., 2005). *Purity* dari *cluster* ke- i adalah:

$$p_i = \max_j p_{ij} \quad (4)$$

Sedangkan *purity* secara keseluruhan adalah:

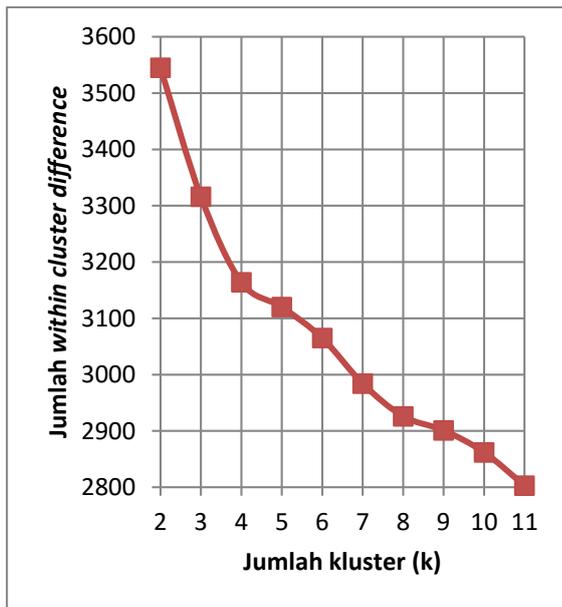
$$purity = \sum_{i=1}^k \frac{m_i}{m} p_i \quad (5)$$

4. HASIL DAN PEMBAHASAN

Clustering dilakukan pada berbagai nilai k ($k = 2, 3, \dots, 10$). Nilai *within cluster difference* terkecil untuk tiap nilai k ditampilkan pada Gambar 2. Berdasarkan kaidah *elbow method*, maka ada dua alternatif k terbaik yaitu $k=4$ dan $k=8$. Ini karena pada $k=4$ dan $k=8$ terdapat bentuk melandai seperti siku (*elbow*). Oleh karena itu, akan dilakukan analisis dua macam nilai k tersebut.

Tabel 6. Hasil clustering dengan k=4

Cluster	Sentroid	Jumlah anggota	Contoh Anggota	Identifikasi Berdasar Sentroid	
				Jenis Masakan	Tipe
A1	bawang merah, bawang putih, cabe merah, garam, gula, kemiri, kunyit, lengkuas, serai, tomat	55	Sayur Asem Banjar Gangan asam kepala patin Gangan Kangkung Keladi Gangan Asam Haruan Ayam Suir Bumbu Habang	sayur asam	spesifik
A2	ayam, bawang merah, bawang putih, cengkeh, daun bawang, garam, gula, jahe, kapulaga, kayu manis, merica, pala, seledri	65	Soto Banjar Soto Banjar khas kalimantan Soto banjar sederhana Bistik ayam banjar	soto banjar	spesifik
A3	bawang merah, bawang putih, garam	140	Terong Bakar Banjar Ayam Masak Habang Garih Batanak Ikan sepat goreng sambal cacapan Ikan Pekasam Khas Banjar Oseng Mandai (kulit cempedak)	masakan gurih lain-lain	umum
A4	garam, gula, santan, telur	95	Amparan Tatak Pisang Hintalu Karuang Lapis India Sarimuka Ketan Kokoleh Bubur randang kuah santan	kue dan bubur manis	umum



Gambar 2. Jumlah *within cluster difference* berdasar jumlah cluster (*k*)

Hasil cluster dengan k=4 (Tabel 5) menghasilkan kelompok dengan jumlah anggota masing-masing 55, 65, 140, dan 95. Kelompok pertama (A1) setelah diidentifikasi secara manual oleh ahli merupakan kelompok yang dominan berupa jenis masakan “sayur asam”. Kelompok kedua (A2) mewakili masakan “soto Banjar”. Kelompok ketiga (A3) berjumlah paling besar 140 resep, tidak mewakili masakan spesifik tetapi diinterpretasi sebagai “masakan asin/gurih” secara umum, karena sentroid yang dihasilkan berupa (bawang merah,

bawang putih, garam). Sedangkan kelompok keempat (A4), juga tidak mewakili satu masakan khusus, melainkan gabungan dan berbagai resep kue manis dan bubur manis dengan sentroid berupa (garam, gula, santan, telur).

Tabel 8. Nilai purity untuk k=4

Cluster	Jenis masakan berdasar sentroid	Jumlah Anggota (m_i)	Benar	Purity
A1	sayur asam	55	32	0,582
A2	soto banjar	65	45	0,692
A3	masakan gurih lain-lain	140	140	1,000
A4	kue dan bubur manis	95	76	0,800
Purity keseluruhan				0,825

Tabel 9. Nilai purity untuk k=8

Cluster	Jenis masakan berdasar sentroid	Jumlah Anggota (m_i)	Benar	Purity
B1	sayur asam	25	24	0,960
B2	soto banjar	54	43	0,796
B3	kue basah	82	59	0,720
B4	masakan gurih lain-lain	117	117	1,000
B5	masak habang	34	26	0,765
B6	bubur manis	13	8	0,615
B7	kuah ketupat	24	13	0,542
B8	masakan gurih asam	6	5	0,833
Purity keseluruhan				0,831

Hasil cluster dengan $k=8$ menghasilkan pengelompokan yang lebih spesifik (Tabel 7). “Sayur asam” dan “Soto Banjar” kembali teridentifikasi sebagai masing-masing satu kelompok (B1 dan B2). Namun ada jenis masakan lain yang teridentifikasi ke dalam kelompok tersendiri yaitu “masak habang” (B5) dan “kuah ketupat” (B7). Selanjutnya ada kelompok yang bersifat umum “masakan gurih asam” (B8), serta “masakan gurih lainnya” (B4).

Masakan bersifat manis yang pada $k=4$ berada pada satu kelompok (A4), namun di $k=8$ terbagi menjadi dua kelompok, yaitu kelompok “kue basah” (B3) dan kelompok “bubur manis” (B6). Namun kedua kelompok ini masih belum menunjukkan jenis masakan yang spesifik.

Selanjutnya dilakukan evaluasi *purity* tiap kelompok pada tiap *cluster*. Setelah tiap kelompok diberi label berdasar sentroidnya, misal A1 diidentifikasi sebagai kelompok sayur asam, maka untuk tiap anggota *cluster* A1 tersebut diperiksa apakah betul tergolong resep sayur asam. Hasil lengkap nilai *purity* tiap *cluster* ditampilkan pada Tabel 8 dan Tabel 9.

Nilai *purity* pada kelompok makanan yang spesifik (seperti sayur asam) secara umum lebih kecil daripada kelompok yang makanan yang umum (seperti masakan gurih). Namun semua nilai *purity* lebih besar daripada 0,5, sehingga dapat dikatakan bahwa deskripsi jenis masakan berdasar sentroid tiap *cluster* bisa mewakili anggota kelompok tersebut.

Tabel 7. Hasil clustering dengan $k=8$

Cluster	Sentroid	Jumlah anggota	Contoh Anggota	Identifikasi Berdasar Sentroid	
				Jenis Masakan	Tipe
B1	bawang merah, bawang putih, cabe merah, garam, gula, kacang panjang, kemiri, kol, kunyit, lengkuas, serai, terasi, timun, tomat	25	Gangan Asam Patin Khas Banjar Sayur asam banjar kepala haruan Sayur Asam Banjar Sayur asam banjarmasin Ikan baung gangan asam	sayur asam	spesifik
B2	ayam, bawang merah, bawang putih, bunga lawang, cengkeh, daun bawang, garam, gula, jahe, kapulaga, kayu manis, merica, pala, seledri	54	Soto banjar Sop Banjar ala Bunda Zio Soto banjar limau nipis Soto Banjar khas kalimantan	soto banjar	spesifik
B3	garam, gula, santan, telur	82	Lempeng Nangka Lempeng Pisang Lapis india Amparan tatak Hintalu karuang	kue basah	umum
B4	bawang merah, bawang putih, garam	117	Ayam panggang banjar & acar Karih ayam banjar Iwak karing khas banjar Ipau/lasagna banjar Mie bancir Pakasam	masakan gurih lain-lain	umum
B5	bawang merah, bawang putih, cabe merah, garam, gula, gula merah, jahe, kayu manis, minyak goreng	34	Hintalu Masak Habang Ayam Masak Habang Ati Ampela Masak Habang Ikan Gabus Masak Habang Ayam Bistik Banjar	masak habang	spesifik
B6	daun pandan, garam, gula, gula merah, kapur sirih, pandan, santan, tepung beras	13	Kokoleh Kokoleh (Bubur Sum-sum Banjar) Babongko banjar	bubur manis	umum
B7	bawang merah, bawang putih, cabe merah, garam, gula, kemiri, kunyit, lengkuas, santan, serai	26	Gangan Asam Daging Gangan tungkul / Sayur Jantung Pisang Garih betanak Gangan asam kepala patin	kuah ketupat	spesifik
B8	bawang merah, bawang putih, cabe rawit, garam, gula, kunyit, serai	6	Pekasam durian Pepes Ikan Kembung Banjar Pedas Gangan asam tulangan Tumis pekasam	masakan gurih asam	umum

5. KESIMPULAN

Telah dilakukan eksperimen *clustering* dengan metode *k*-modes terhadap data bahan masakan Banjar yang bersumber dari suatu *recipe sharing website*. Untuk kasus ini, ada 2 alternatif jumlah *cluster* yang ideal, yaitu $k=4$ dan $k=8$. Jumlah *cluster* 4 menghasilkan kelompok yang lebih umum, namun teridentifikasi bahwa soto Banjar dan sayur asam adalah dua kelompok yang populer. Sedangkan jumlah *cluster* 8 menghasilkan kelompok yang lebih sempit. Di samping soto Banjar dan sayur asam, juga bisa mengidentifikasi secara khusus jenis kelompok "masak *habang*" dan "kuah ketupat". Selain jenis masakan yang telah disebutkan, *k-modes* hanya mengidentifikasi kelompok yang umum seperti "masakan gurih" atau "kue basah".

Pada penelitian ini, fitur yang digunakan terbatas berupa bahan makanan. Perlu diteliti pemodelan data resep menggunakan fitur lain, misal TF-IDF. Selain itu, bagian judul dan instruksi memasak dari suatu resep juga bisa diekstraksi untuk membangun fitur.

Pola kelompok yang dihasilkan mewakili pola khusus dari resep masakan Banjar saja. Selanjutnya bisa digali pola masakan dari berbagai daerah lain. Dengan demikian, bisa didapat gambaran yang lebih lengkap mengenai persamaan dan perbedaan bahan pada masakan yang populer di tiap daerah.

6. DAFTAR PUSTAKA

- AHN, Y., AHNERT, S. E., BAGROW, J. P., & BARABÁSI, A. 2011. Flavor Network And The Principles Of Food Pairing. *Sci. Rep. 1*. doi:10.1038/srep00196.
- GOUTTE, C., TOFT, P., ROSTRUP, E., NIELSEN, F. A., & HANSEN, L. K. 1999. On Clustering fMRI Time Series. *NeuroImage*. 9 (3): 298–310. doi:10.1006/nimg.1998.0391.
- HUANG, Z. 1998. Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Mining and Knowledge Discovery*, 2(3), 283–304.
- JAIN, ANUPAM, RAKHI N K, & BAGLER, G.. 2015. Spices Form The Basis Of Food Pairing In Indian Cuisine. Arxiv.Org.
- RUFAIDA, S. 2016. Klasifikasi Resep Makanan Pendamping Air Susu Ibu (MPASI) Dengan Metode Decision Tree ID3 Berdasarkan Bahan Makanan. Skripsi, Universitas Sebelas Maret.
- SBS FOOD,. 2014. About Indonesian Food. *Food*. Diakses 20/05/2016.
- STEINBACH, M., KARYPIS, G. & KUMAR, V. 2000. A comparison of document clustering techniques. *KDD workshop on text mining* (Vol. 400, No. 1, pp. 525-526).
- SU, HAN, SHAN, M. K., LIN, T. W., CHANG, J., & LI, C. T. 2014. Automatic Recipe Cuisine Classification By Ingredients. *Proceedings Of The 2014 ACM International Joint Conference On Pervasive And Ubiquitous Computing Adjunct Publication - Ubicomp '14 Adjunct*. doi:10.1145/2638728.2641335.
- TAN, P.-N., STEINBACH, M., & KUMAR, V. 2005. Introduction to Data Mining. *Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA*.
- ZHU, YU-XIAO, HUANG, J., ZHANG, Z.K., ZHANG, Q. M, ZHOU, T., & AHN, Y. Y. 2013. Geography And Similarity Of Regional Cuisines In China. *Plos ONE* 8 (11): e79161. doi:10.1371/journal.pone.0079161.