



Pola Asosiasi Bahan pada Resep Masakan Daerah dengan Algoritma Apriori

Fatma Indriani^a

^aProgram Studi Ilmu Komputer, Fakultas MIPA, Universitas Lambung Mangkurat, f.indriani@unlam.ac.id

Abstract

Recipe sharing websites are popular among internet users. Recipes data can be used to mine insight about food culture in the community. Indonesia has a rich cuisine tradition. There are more than 6000 islands in Indonesia, and every region has its own cuisine characteristic. One characteristic is the combination of ingredients and spices that are often used in the region. These common ingredient combinations can be found using association analysis from online recipe sites. This paper presented the process of data collection of Banjar cuisine from a popular recipe sharing website, cookpad.com, followed by association analysis of recipe ingredients using the Apriori algorithm.

Keywords: web mining, recipe, Banjar, association rule, Apriori

Abstrak

Website untuk *sharing* resep digunakan oleh berbagai pengguna internet untuk saling berbagi resep. Data resep tersebut bisa dimanfaatkan untuk menggali informasi mengenai berbagai makanan yang populer di masyarakat. Indonesia memiliki tradisi masakan yang kaya di dunia. Ada lebih dari 6000 pulau di Indonesia, dan hampir tiap daerah memiliki makanan khas. Salah satu ciri khas yaitu kombinasi bahan dan bumbu yang sering digunakan pada daerah tersebut. Kombinasi bahan tersebut dapat dicari dari data resep di website menggunakan teknik analisis asosiasi. Pada makalah ini dipaparkan proses pengambilan data masakan daerah Banjar dari website Cookpad, kemudian dilakukan identifikasi pola asosiasi bahan resep dengan menggunakan algoritma Apriori. Nilai minimum support yang ideal untuk data resep ini adalah 0,10, sedangkan nilai confidence yaitu 0,9. Pola asosiasi yang dihasilkan ada 9 buah, meliputi bahan-bahan ayam, kapulaga, cengkeh, jahe, kayu manis, dan pala dengan kombinasi yang berbeda.

Kata kunci: web mining, resep, Banjar, aturan asosiasi, Apriori

© 2017 Prosiding SISFOTEK

1. Pendahuluan

Makanan adalah sumber energi sekaligus kebutuhan pokok dari manusia. Zhu, dkk [1] menyatakan bahwa daerah yang dekat secara geografis juga memiliki kemiripan kuliner yang erat. Namun Indonesia memiliki tradisi masakan yang kaya di dunia [2]. Ada lebih dari 6000 pulau di Indonesia, dan hampir tiap daerah memiliki makanan khas.

Salah satu ciri khas yaitu kombinasi bahan/bumbu yang sering digunakan pada daerah tersebut. Pengetahuan mengenai kombinasi bahan ini bisa ditanyakan pada penduduk lokal atau ahli masak lokal, namun untuk suatu daerah yang berlokasi jauh, akan dialami kesulitan menemukan ahli yang tepat dan relevan. Di lain pihak, resep berbagai masakan khas baik daerah Indonesia maupun di dunia tersedia secara bebas dan

banyak di internet dan bisa digali menggunakan teknik data mining dan text mining.

Salah satu teknik penggalian pola hubungan adalah analisis aturan asosiasi. Pada awalnya, teknik aturan asosiasi diterapkan untuk menganalisis transaksi keranjang belanja untuk melihat pola kelompok barang apa yang sering dibeli bersamaan. Teknik pencarian aturan asosiasi juga bisa diterapkan untuk dataset jenis lain. Pada daftar bahan suatu resep, teknik asosiasi akan menemukan kombinasi bahan apa yang sering muncul dalam sekelompok resep. Jika sekelompok resep ini dibatasi pada masakan dari daerah tertentu, misal masakan khas Banjar, maka pola yang ditemukan bisa mendeskripsikan ciri khas masakan daerah tersebut.

Selanjutnya pada makalah ini akan dipaparkan tinjauan terdahulu terkait analisis resep masakan (Bagian II). Bagian III akan menjelaskan metodologi yang

digunakan meliputi pengumpulan data resep masakan daerah Banjar, prapemrosesan data, serta pengolahan data dengan metode Apriori. Bagian IV menampilkan hasil dan pembahasan. Bagian V menyampaikan kesimpulan dan saran.

2. Tinjauan Pustaka

Wang, dkk [3] merepresentasikan resep masakan China secara semantik dalam bentuk graf untuk mengoptimalkan hasil pencarian resep berdasar query pengguna. Tidak ada proses mining atau learning pada penelitian ini karena pembuatan representasi graf tiap resep dilakukan secara manual. Wang dkk berfokus pada proses retrieval, di mana resep yang mirip bahan dasar atau tahap masakannya, bisa dideteksi walaupun judul dan bahannya banyak perbedaan.

Dalam ranah data mining, penelitian dalam pengolahan data resep melibatkan *teknik supervised* dan *unsupervised learning*. Penerapan teknik *supervised learning* dilakukan oleh Su, dkk [4], 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 [5] yang membuat model klasifikasi resep MPASI (Makanan Pendamping ASI) ke dalam target kelompok umur, dengan metode ID3.

Sedangkan penelitian yang menerapkan teknik *unsupervised learning* yaitu Jain dkk [6]. 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 [7].

Nadamoto dkk [8] 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 *search* ke dalam kelompok-kelompok yang mirip.

Pada makalah ini, algoritma apriori digunakan untuk meneliti resep masakan daerah berbahasa Indonesia, memiliki kemiripan dengan penelitian [6], karena sama-sama menggunakan teknik *unsupervised* pada bahan resep. Namun Jain dkk [6] hanya meneliti

pasangan bahan (hubungan antara dua bahan), sedangkan dengan algoritma apriori, ada kemungkinan menemukan asosiasi yang melibatkan lebih dari dua bahan. Ini akan memberikan informasi yang lebih kaya mengenai ciri khas suatu masakan daerah.

3. Metodologi Penelitian

Langkah-langkah yang dilakukan terdiri dari beberapa tahap sebagai berikut.

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:

- a. Ekstraksi bahan (*ingredient*). Hanya informasi bahan yang digunakan pada penelitian ini. Informasi mengenai langkah-langkah memasak 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</p> <p>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</p> <p>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>

b. *Cleaning* data meliputi pembuangan karakter non huruf, pembuangan *stopwords* yang dimodifikasi untuk data resep, serta penyeragaman ejaan dan istilah. Jumlah variabel bahan akhir yaitu sebanyak 449 bahan unik, dan rata-rata tiap resep mengandung 12,23 macam bahan. Contoh hasil pada tahap ini dapat dilihat di tabel berikut.

Tabel 2. Contoh Hasil *Cleaning*

<p>Sayur/Gangan/Jangan Asem khas Banjar https://cookpad.com/id/resep/2144448-sayurganganjangan-asem-khas-banjar</p> <p>patin;timun;kol;tomat;cabe;bawang merah;kemiri;lengkuas;kunyit;serai;cabe rawit;cabe;garam;gula;saus tiram;panic</p>
<p>Soto Banjar https://cookpad.com/id/resep/2135070-soto-banjar</p> <p>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;telur bebek;ayam</p>

c. Vektorisasi data menghasilkan matriks biner berukuran 355 x 449

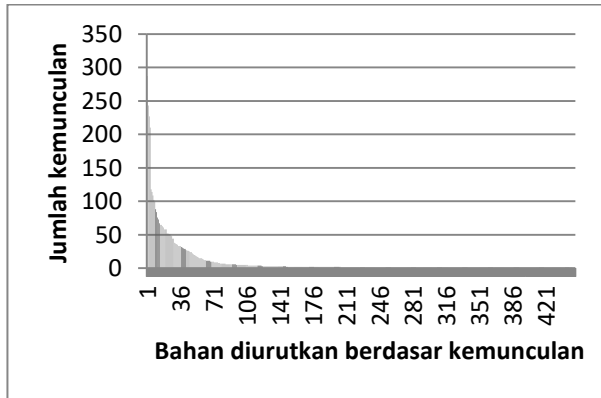
3.3 Pencarian aturan asosiasi

Apriori adalah algoritma dasar untuk menemukan frequent itemset yang diperkenalkan oleh Agrawal dan Srikant [9]. Algoritma ini menggunakan prinsip apriori, yaitu “Jika suatu itemset adalah frequent, maka semua subsetnya juga frequent.” Prinsip ini digunakan untuk memangkas kandidat itemset.

Pada tahap ini, dicari *frequent itemset* serta aturan asosiasi dari data bahan resep menggunakan algoritma apriori. Pertama akan dicari pengaruh minimum support dan minimum confidence terhadap rule. Selanjutnya akan dipaparkan aturan asosiasi yang ditemukan beserta evaluasi berdasar nilai *support*, *confidence*, dan *lift*.

4. Hasil dan Pembahasan

Dari 449 bahan yang ada, masing-masing memiliki kemunculan yang sangat bervariasi, dengan rata-rata kemunculan 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. Frekuensi Kemunculan Bahan

Tabel 3. 20 Bahan Dengan Kemunculan Paling Sering

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	terasi	58

Agar pola yang dihasilkan lebih bermakna (*useful*), maka bahan yang kemunculannya lebih dari 50% akan diabaikan, yaitu *garam*, *bawang merah*, *bawang putih*, dan *gula*, sehingga jumlah bahan menjadi 445.

4.1 Pengaruh Minimum Support terhadap Frequent itemset

Untuk melihat pengaruh minimum support terhadap frequent 1-itemset, digunakan 4 nilai support yaitu 5%, 8%, 10%, dan 12% (Tabel 4). Semakin rendah nilai min-support, maka semakin banyak bahan yang lolos pada frequent 1-itemset. Min-support yang rendah juga akan meningkatkan frequent itemset pada level yang lebih tinggi. Sebaliknya semakin tinggi min support, maka akan semakin sedikit variasi bahan. Hal ini nanti akan berdampak pada pola yang dihasilkan. Jika ingin mencari pola yang lebih umum, maka nilai support sebaiknya ditinggikan.

Tabel 4. Jumlah Frequent Itemset berdasar Minimum Support

Min Sup	Min Count	Jumlah Frequent itemset					
		1-item set	2-item set	3-item set	4-item set	5-item set	6-item set
0,05	18	49	155	177	130	59	13
0,08	28	37	68	39	10	0	0
0,10	36	28	31	8	0	0	0
0,12	43	25	16	2	0	0	0

4.2 Pengaruh Minimum Support dan Minimum Confidence terhadap Jumlah Rule

Percobaan selanjutnya yaitu mengkombinasikan beberapa nilai support dan confidence. Untuk support, digunakan 4 macam nilai yaitu 0,05; 0,08; 0,10; dan 0,12. Sedangkan untuk minimum confidence digunakan nilai 0,7; 0,8; 0,9; dan 0,95. Dilihat dari jumlah aturan asosiasi yang ditemukan (Tabel 5), maka kombinasi minimum support yang rendah dan minimum confidence yang juga rendah akan menghasilkan rule yang sangat banyak. Sedangkan jika support tinggi dan confidence tinggi, maka jumlah rule akan menjadi sedikit, bahkan tidak ada. Rule yang terlalu banyak akan mengakibatkan pola yang sulit dianalisis dan kurang bermanfaat. Oleh karena itu, peneliti menyarankan untuk kasus ini, minimum support dan minimum confidence yang tepat yaitu masing-masing 0,1 dan 0,9, menghasilkan 9 buah aturan asosiasi.

Tabel 5. Jumlah Rule Asosiasi berdasar nilai Min-Support dan Min-Confidence

Jumlah Rule Asosiasi		Minimum Support			
		0,05	0,08	0,10	0,12
Minimum Confidence	0,70	1926	175	37	13
	0,80	1135	101	19	6
	0,90	573	43	9	2
	0,95	242	7	-	-

4.3 Pola Asosiasi dengan Minimum Support 0,1 dan Minimum Confidence 0,9

Berdasar pemaparan sebelumnya, maka digunakan parameter minimum support 0,1 dan minimum confidence 0,9. Jumlah 1, 2, 3 frequent itemset yang ditemukan sebanyak 28, 31, dan 8 macam (Tabel 6). Namun rule asosiasi yang ditemukan hanya berjumlah 9 buah (Tabel 7). Berdasar pengamatan, 9 rule asosiasi tersebut hanya meliputi 6 bahan, yaitu ayam, kapulaga, cengkeh, jahe, kayu manis, dan pala. Keenam bahan ini adalah bahan yang umum digunakan untuk membuat soto banjar.

Tabel 6. Frequent Itemset

Frequent Itemset	Jumlah	Bahan (Count)
1-Itemset	28	asam jawa (51) ayam (84) bawang goreng (52) cabe merah (99) cabe rawit (52) cengkeh (65) daun bawang (49) gula merah (102)

		jahe (114) jeruk nipis (36) kapulaga (44) kayu manis (88) kecap manis (37) kemiri (76) kentang (44) kunyit (73) lengkuas (48) merica (58) minyak goreng (67) pala (51) penyedap (58) santan (118) seledri (61) serai (65) telur (109) terasi (58) terigu (38) tomat (63)
2-Itemset	31	ayam, cengkeh (50) ayam, jahe (56) ayam, kapulaga (39) ayam, kayu manis (55) ayam, merica (43) ayam, pala (43) ayam, seledri (45) cabe merah, jahe (43) cabe merah, kemiri (36) cabe merah, kunyit (39) cabe merah, tomat (38) cengkeh, jahe (47) cengkeh, kapulaga (41) cengkeh, kayu manis (58) cengkeh, pala (43) cengkeh, seledri (37) gula merah, jahe (45) gula merah, kayu manis (38) gula merah, santan (44) jahe, kayu manis (65) jahe, kemiri (42) jahe, pala (37) jahe, seledri (37) kapulaga, kayu manis (40) kayu manis, pala (40) kayu manis, seledri (39) kemiri, kunyit (44) kemiri, serai (39) kunyit, lengkuas (36) kunyit, serai (43) santan, telur (51)
3-Itemset	8	ayam, cengkeh, jahe (37) ayam, cengkeh, kapulaga (37) ayam, cengkeh, kayu manis (45) ayam, cengkeh, pala (36) ayam, jahe, kayu manis (41) cengkeh, jahe, kayu manis (44) cengkeh, kapulaga, kayu manis (37) cengkeh, kayu manis, pala (37)

Tabel 7. Hasil Pola Asosiasi

No	Rule	Support	Confidence	Lift
1	ayam, kapulaga => cengkeh	0,11	0,95	5,18
2	cengkeh, jahe => kayu manis	0,13	0,94	3,78
3	kapulaga => cengkeh	0,12	0,93	3,78
4	kapulaga, kayu manis => cengkeh	0,11	0,93	5,05
5	kayu manis, pala => cengkeh	0,11	0,93	5,05
6	kapulaga => kayu manis	0,12	0,91	3,67

7	cengkeh, kapulaga => ayam	0,12	0,90	3,81
8	cengkeh, kapulaga => kayu manis	0,10	0,90	3,64
9	ayam, cengkeh => kayu manis	0,14	0,90	3,63

5. Kesimpulan

5.1 Simpulan

Dari pemaparan di atas, telah dilakukan identifikasi pola asosiasi bahan resep masakan Banjar yang digali dari website *sharing* resep. Nilai minimum support yang ideal untuk data resep ini adalah 0,10, sedangkan nilai confidence yang ideal yaitu 0,9 agar menghasilkan pola yang tidak terlalu banyak. Pola yang dihasilkan ada 9 buah, meliputi bahan-bahan ayam, kapulaga, cengkeh, jahe, kayu manis, dan pala dengan kombinasi yang berbeda.

5.2 Saran

Adapun saran untuk penelitian selanjutnya adalah:

1. Pola asosiasi cenderung menemukan pola dari jenis masakan yang sering dibagikan, dalam hal ini jenis masakan gurih. Jenis masakan lain yang lebih sedikit seperti kue-kuean tidak muncul pola asosiasinya, karena nilai *support*-nya tidak mencukupi. Sebaiknya teliti metode clustering atau metode lainnya untuk menggali pola lebih lanjut.
2. Pola asosiasi yang dihasilkan mewakili pola khusus dari daerah Banjar saja. Sebaiknya dilanjutkan dengan menggali pola masakan dari berbagai daerah lain. Dengan demikian,

akan didapat gambaran yang lebih lengkap mengenai persamaan dan perbedaan bahan pada masakan yang populer di tiap daerah.

6. Daftar Rujukan

- [1] Zhu, Yu-Xiao, Junming Huang, Zi-Ke Zhang, Qian-Ming Zhang, Tao Zhou, dan Yong-Yeol Ahn. 2013. "Geography And Similarity Of Regional Cuisines In China". *Plos ONE* 8 (11): e79161. doi:10.1371/journal.pone.0079161.
- [2] SBS Food. 2014. "About Indonesian Food". *Food*. Diakses 20/05/2016
- [3] Wang, Liping, Qing Li, Na Li, Guozhu Dong, dan Yu Yang. 2008. "Substructure Similarity Measurement In Chinese Recipes". *Proceeding Of The 17Th International Conference On World Wide Web - WWW '08*. doi:10.1145/1367497.1367629.
- [4] Su, Han, Man-Kwan Shan, Ting-Wei Lin, Janet Chang, dan Cheng-Te Li. 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.
- [5] Rufaida, Solikhah. 2016. "Klasifikasi Resep Makanan Pendamping Air Susu Ibu (MPASI) Dengan Metode Decision Tree ID3 Berdasarkan Bahan Makanan". Skripsi, Universitas Sebelas Maret.
- [6] Jain, Anupam, Rakhi N K, dan Ganesh Bagler. 2015. "Spices Form The Basis Of Food Pairing In Indian Cuisine". Arxiv.Org.
- [7] Ahn, Yong-Yeol, Sebastian E. Ahnert, James P. Bagrow, dan Albert-László Barabási. 2011. "Flavor Network And The Principles Of Food Pairing". *Sci. Rep. 1*. doi:10.1038/srep00196.
- [8] Nadamoto, A, Hanai, S., and Nanba, H., "Clustering for Similar Recipes in User-Generated Recipe Sites Based on Main Ingredients and Main Seasoning", In *2016 19th International Conference on Network-Based Information Systems (NBIS)*, 2016.
- [9] Agrawal, R. dan Srikant, R., 1994. "Fast algorithms for mining association rules." *Proc. 20th int. conf. very large data bases, VLDB* (Vol. 1215, pp. 487-499).