

# SKRIPSI

## KLASTERISASI RESEP MAKANAN BERDASARKAN BAHAN YANG DIGUNAKAN DAN KOMPOSISI BAHAN DI MAKANAN TERSEBUT



Denise Stevani Gebriella

NPM: 2017730045

PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS  
UNIVERSITAS KATOLIK PARAHYANGAN  
2022

**UNDERGRADUATE THESIS**

**CLUSTERING OF FOOD RECIPES BASED ON THE  
INGREDIENTS USED AND THE COMPOSITION OF THE  
INGREDIENTS IN THE FOOD**



**Denise Stevani Gebriella**

**NPM: 2017730045**

**DEPARTMENT OF INFORMATICS  
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES  
PARAHYANGAN CATHOLIC UNIVERSITY  
2022**

# LEMBAR PENGESAHAN

## KLASTERISASI RESEP MAKANAN BERDASARKAN BAHAN YANG DIGUNAKAN DAN KOMPOSISI BAHAN DI MAKANAN TERSEBUT

Denise Stevani Gebriella

NPM: 2017730045

Bandung, 28 Juni 2022

Menyetujui,

Pembimbing

Digitally signed  
by Luciana  
Abednego

Luciana Abednego, M.T.

Ketua Tim Penguji

Digitally signed  
by Mariskha Tri  
Adithia

Mariskha Tri Adithia, P.D.Eng

Anggota Tim Penguji

Digitally signed  
by Maria V.  
Claudia M.

Maria Veronica, M.T.

Mengetahui,

Ketua Program Studi

Digitally signed  
by Mariskha Tri  
Adithia

Mariskha Tri Adithia, P.D.Eng

## PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

### **KLASTERISASI RESEP MAKANAN BERDASARKAN BAHAN YANG DIGUNAKAN DAN KOMPOSISI BAHAN DI MAKANAN TERSEBUT**

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,  
Tanggal 28 Juni 2022



Denise Stevani Gebriella  
NPM: 2017730045

## ABSTRAK

Di era digital ini, resep makanan Indonesia semakin mudah ditemukan. Resep bisa didapatkan dari banyak tempat, salah satunya adalah dari situs resep makanan di internet. Resep merupakan tata cara pembuatan makanan yang berisi nama dari resep, bahan dan lauk yang digunakan, dan tata cara pembuatannya. Jika pengguna melihat sebuah resep di sebuah laman di internet, ada kemungkinan pengguna tersebut ingin melihat resep yang sejenis atau mirip dengan resep sebelumnya. Untuk mencari resep yang sejenis atau mirip dibutuhkan pengelompokan resep. Untuk mengelompokkan resep dibutuhkan teknik pengelompokan yang dinamakan klasterisasi. Klasterisasi adalah teknik yang mengelompokkan objek-objek dengan karakteristik yang serupa ke dalam satu klaster. Ada beberapa cara untuk melakukan klasterisasi di antaranya adalah *k-means* dan *agglomerative hierarchical*. Setiap resep akan direpresentasikan dengan sebuah vektor yang dimensinya adalah komposisi bahan di resep tersebut. Untuk dapat menghitung kemiripan antar resep maka dibutuhkan perhitungan jarak. Digunakan dua buah perhitungan yaitu perhitungan *cosine similarity* dan *Euclidean distance*.

Pada penelitian ini, dibangun sebuah perangkat lunak yang dapat mengelompokkan kumpulan resep makanan Indonesia yang berisi nama, bahan, dan kuantitas dari setiap bahan yang digunakan pada resep. Perangkat lunak mengimplementasikan metode *k-means* dan *agglomerative hierarchical*. Tujuan dari penelitian ini adalah untuk melakukan klasterisasi dan menganalisis hasil klasterisasi untuk resep makanan Indonesia. Hasil dari klasterisasi akan dianalisis untuk menentukan apakah resep-resep yang berada di klaster yang sama memiliki kemiripan dalam hal cita rasa, atau jenis makanan. Analisa juga dilakukan untuk daerah asal resep pada setiap resep di klaster, apakah daerah terbagi dengan baik. Analisis akan dilakukan pada klaster yang memiliki *intracluster* terkecil atau terbaik. Pengujian juga dilakukan untuk melihat perbandingan hasil yang lebih baik berdasarkan algoritma penggunaan *k-means* atau *agglomerative hierarchical*, penggunaan perhitungan vektor *cosine* atau *Euclidean*, dan penggunaan vektor fitur bahan tanpa lauk atau bahan dengan lauk.

Hasil pengujian menunjukkan bahwa makanan Indonesia bisa dikelompokkan. Berdasarkan hasil dari klasterisasi tidak begitu terlihat ciri khas dari sebuah klaster dan tidak terlalu jelas perbedaan antara suatu klaster dengan klaster yang lain. Hasil ini dapat terjadi karena pada dataset yang digunakan banyak makanan Indonesia yang menggunakan bahan dan bumbu yang sama, sehingga tidak terlalu terlihat perbedaan antar satu kelompok dan kelompok lainnya. Makanan di setiap daerah Indonesia pun memiliki jenis makanan yang berbeda. Hal ini menyebabkan satu daerah bisa berada dalam kelompok yang berbeda.

**Kata-kata kunci:** klasterisasi resep makanan Indonesia, *k-means*, *agglomerative hierarchical*

## ABSTRACT

In this digital era, Indonesian food recipes are easier to find. Recipes can be obtained from many places, one of which is from food recipe sites on the internet. A recipe is a procedure for making food that contains the name of the recipe, the ingredients and side dishes used, and the procedure for making it. If a user sees a recipe on a page on the internet, it is possible that the user wants to see a recipe that is similar to the previous recipe. To find similar recipes, it takes grouping of recipes. To group recipes, it takes a grouping technique called clustering. Clustering is a technique that groups objects with similar characteristics into one cluster. There are several ways to do clustering including k-means and agglomerative hierarchical. Each recipe will be represented by a vector whose element is the composition of the ingredients in the recipe. So to calculate the similarity between recipes, it is necessary to calculate the distance. Two calculations are used, namely the calculation of cosine similarity and Euclidean distance. The purpose of this study is to perform clustering and analyze the results of clustering for Indonesian food recipes.

In this study, a software was built that can group a collection of Indonesian food recipes containing the name, ingredient, and quantity of each ingredient used in the recipe. The software implements the k-means and agglomerative hierarchical methods. The results of the clustering will be analyzed to determine whether the recipes in the same cluster have similarities in terms of taste, or type of food. Analysis was also carried out for the area of origin of the prescription for each recipe in the cluster, whether the area was well divided or not. The analysis will be carried out on the cluster that has the smallest or best intracluster. Tests were also carried out to see a better comparison of results based on the use of k-means or agglomerative hierarchical algorithms, the use of cosine or Euclidean vector calculations, and the use of feature vectors of ingredients without side dishes or ingredients with side dishes.

The test results show that Indonesian food can be grouped but based on the results of the clustering, the characteristics of a cluster are not very visible and the differences between one cluster and another are not very clear. This result can happen because the dataset used a lot of Indonesian food that uses the same ingredients and seasonings, so there is not much difference between one group and another. Food in each region of Indonesia also has different types of food. This causes one area to be in a different group.

**Keywords:** Clustering Indonesian food recipes, k-means, agglomerative hierarchical

## KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas kasih dan karunia-Nya, penulis dapat menyelesaikan penyusunan skripsi yang berjudul "Klasterisasi Resep Makanan Berdasarkan Bahan Yang Digunakan dan Komposisi Bahan Di Makanan Tersebut". Penulis menyadari bahwa penulis tidak luput dari kesalahan dan saat penulisan skripsi ini penulis menghadapi banyak kendala dan berbagai masalah. Dibalik penyusunan skripsi ini, penulis sadar bahwa penyusunan skripsi ini tidak mungkin dilakukan tanpa dukungan dan bantuan orang-orang terdekat. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih kepada: Secara khusus, penulis ingin berterima kasih kepada:

1. Tuhan Yang Maha Esa atas Kasih, Berkah dan Rahmat-Nya.
2. Diri sendiri.
3. Orang tua beserta kakak laki-laki penulis yang selalu memberikan dukungan kepada penulis baik berupa doa atau dukungan mental serta materil.
4. Keluarga besar penulis yang selalu memberikan dukungan mental.
5. Kucing dan Anjing penulis yang selalu menemani penulis.
6. Ibu Natalia, M.Si. selaku dosen pembimbing yang telah membimbing penulis dan memberikan dukungan dan bantuan kepada penulis dalam proses penyusunan skripsi ini.
7. Ibu Mariskha Tri Adithia, P.D.Eng dan Ibu Maria Veronica, M.T. selaku penguji yang telah memberikan kritik dan saran yang membangun.
8. Sahabat perempuan informatika angkatan 17 seperti Friska Christiana, Melody Victorian, Hendrika Valeska, Cristine Artanty, Linna Trisnawati, dan Rachel Florencia yang sudah memberikan dukungan, bantuan, dan semangat kepada penulis.
9. Sahabat laki-laki informatika angkatan 17 seperti Michael Liondy, Juan Nandriisa, dan Enrico Wibawa yang juga sudah memberikan dukungan, bantuan, dan semangat kepada penulis.
10. Teman-teman informatika angkatan 2017 yang sudah sama-sama berbagi ilmu dan dukungan.

Penulis menyadari bahwa penelitian ini masih jauh dari kata sempurna. Oleh karena itu, penulis memohon maaf jika terdapat kesalahan. Penulis juga mengharapkan kritik dan saran yang membangun untuk menyempurnakan penelitian ini. Semoga penelitian ini dapat bermanfaat bagi pembaca atau menjadi inspirasi untuk penelitian-penelitian berikutnya.

Bandung, Juni 2022

Penulis

# DAFTAR ISI

<b>KATA PENGANTAR</b>	<b>xv</b>
<b>DAFTAR ISI</b>	<b>xvii</b>
<b>DAFTAR GAMBAR</b>	<b>xix</b>
<b>DAFTAR TABEL</b>	<b>xxi</b>
<b>DAFTAR KODE PROGRAM</b>	<b>xxv</b>
<b>1 PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang . . . . .	1
1.2 Rumusan Masalah . . . . .	2
1.3 Tujuan . . . . .	2
1.4 Batasan Masalah . . . . .	2
1.5 Metodologi . . . . .	3
1.6 Sistematika Pembahasan . . . . .	3
<b>2 LANDASAN TEORI</b>	<b>5</b>
2.1 Web Scraping [1] . . . . .	5
2.2 Data Mining [2] . . . . .	6
2.3 Normalisasi Data[3] . . . . .	7
2.4 Kemiripan antar Vektor [2] . . . . .	8
2.5 Klasterisasi [4] . . . . .	8
2.5.1 Agglomerative Hierarchical [4] . . . . .	9
2.5.2 K-Means [4] . . . . .	10
2.6 Intracluster [2] . . . . .	11
2.7 Resep[5] . . . . .	12
<b>3 ANALISIS PERANGKAT LUNAK</b>	<b>13</b>
3.1 Analisis Masalah . . . . .	13
3.2 Analisis Dataset . . . . .	17
3.3 Studi Kasus . . . . .	21
3.4 Gambaran Umum Perangkat Lunak . . . . .	27
3.4.1 Diagram Aktivitas . . . . .	27
3.4.2 Diagram Use Case . . . . .	28
3.4.3 Analisis Diagram Use Case . . . . .	29
<b>4 PERANCANGAN PERANGKAT LUNAK</b>	<b>33</b>
4.1 Diagram Kelas Clustering Lengkap . . . . .	33
4.1.1 Kelas Main . . . . .	33
4.1.2 Kelas Algorithm . . . . .	33
4.1.3 Kelas KMeans . . . . .	34
4.1.4 Hierarchical . . . . .	35



4.2	Perancangan Antarmuka . . . . .	37
<b>5</b>	<b>IMPLEMENTASI DAN PENGUJIAN</b>	<b>39</b>
5.1	Implementasi Antarmuka . . . . .	39
5.2	Pengujian . . . . .	43
5.2.1	Pengujian Fungsional . . . . .	43
5.2.2	Pengujian Eksperimental . . . . .	44
5.3	Kesimpulan Pengujian . . . . .	59
<b>6</b>	<b>KESIMPULAN DAN SARAN</b>	<b>61</b>
6.1	Kesimpulan . . . . .	61
6.2	Saran . . . . .	61
	<b>DAFTAR REFERENSI</b>	<b>63</b>
	<b>A KODE PROGRAM</b>	<b>65</b>
	<b>B HASIL EKSPERIMEN</b>	<b>73</b>
B.1	Tabel hasil Agglomerative Hierarchical Clustering . . . . .	75
B.1.1	Tabel k=4 dengan Cosine Distance, dan Fitur Bahan Tanpa Lauk . . . . .	75
B.1.2	Tabel k=3 dengan Euclidean Distance, dan Fitur Bahan Tanpa Lauk . . . . .	82
B.1.3	Tabel k=6 dengan Cosine Distance, dan Fitur Bahan Dengan Lauk . . . . .	89
B.1.4	Tabel k=4 dengan Euclidean Distance, dan Fitur Bahan Dengan Lauk . . . . .	97
B.2	Tabel hasil K-Means Clustering . . . . .	106
B.2.1	Tabel k=7 dengan Cosine Distance, dan Fitur Bahan Tanpa Lauk . . . . .	106
B.2.2	Tabel k=5 dengan Euclidean Distance, dan Fitur Bahan Tanpa Lauk . . . . .	115
B.2.3	Tabel k=4 dengan Cosine Distance, dan Fitur Bahan Dengan Lauk . . . . .	123

## DAFTAR GAMBAR

2.1	Hierarchical clustering . . . . .	9
2.2	K-means . . . . .	11
3.1	Diagram alur untuk analisa masalah . . . . .	13
3.2	Diagram Alur untuk metode K-Means . . . . .	15
3.3	Diagram alur untuk metode <i>agglomerative hierarchical</i> . . . . .	16
3.4	Hasil data yang diambil dengan <i>web scraping</i> . . . . .	18
3.5	Contoh Dataset yang berisi nama resep dan bahan yang digunakan . . . . .	19
3.6	Contoh Dataset yang sudah terisi lengkap . . . . .	20
3.7	Contoh Dataset yang sudah dinormalisasi . . . . .	20
3.8	Diagram Aktivitas . . . . .	27
3.9	Diagram Use Case . . . . .	28
4.1	Diagram kelas perangkat lunak clustering . . . . .	33
4.2	Rancangan Antarmuka saat menjalankan program . . . . .	37
4.3	Rancangan antarmuka saat program selesai berjalan . . . . .	38
4.4	Rancangan antarmuka saat program mengeluarkan hasil <i>clustering</i> . . . . .	38
4.5	Rancangan Antarmuka saat masukan tidak valid . . . . .	38
5.1	Tampilan perangkat lunak saat dijalankan . . . . .	39
5.2	Tampilan direktori jika pengguna sudah memasukkan <i>file</i> . . . . .	40
5.3	Tampilan saat perangkat lunak menerima masukan dari pengguna . . . . .	40
5.6	Tampilan jika masukan valid . . . . .	42
5.7	Tampilan keluaran klasterisasi . . . . .	42
5.8	Tampilan keluaran <i>dummy</i> data <i>K-Means</i> . . . . .	43
5.9	Tampilan keluaran <i>dummy</i> data <i>Agglomerative Hierarchical</i> . . . . .	44

## DAFTAR TABEL

2.1	Tabel data contoh untuk <i>linkage</i> . . . . .	9
2.2	Hasil dari perhitungan <i>Euclidean distance</i> . . . . .	10
2.3	Hasil dari perhitungan jarak terkecil untuk <i>single linkage</i> . . . . .	10
2.4	Hasil dari perhitungan jarak terkecil untuk <i>single linkage</i> . . . . .	10
3.1	Contoh pengisian kuantitas bahan pada resep . . . . .	20
3.2	Dummy data yang akan digunakan . . . . .	21
3.3	Hasil perhitungan jarak dengan <i>centroid</i> random . . . . .	22
3.4	Hasil perhitungan jarak dengan <i>centroid</i> selanjutnya . . . . .	22
3.5	Hasil perhitungan manual <i>k-means</i> . . . . .	23
3.6	Perhitungan jarak terkecil antar objek ke objek . . . . .	24
3.7	Hasil Perhitungan jarak terkecil antar kelompok ke objek atau objek ke objek dan mengambil yang terkecil . . . . .	24
3.8	Hasil Perhitungan jarak terkecil antar kelompok ke objek atau objek ke objek dan mengambil yang terkecil (2) . . . . .	24
3.9	Hasil perhitungan jarak terkecil antar kelompok ke objek atau objek ke objek dan mengambil yang terkecil (3) . . . . .	25
3.10	Hasil perhitungan jarak terkecil antar kelompok ke objek atau objek ke objek dan mengambil yang terkecil (4) . . . . .	25
3.11	Hasil perhitungan jarak terkecil antar kelompok ke objek atau objek ke objek dan mengambil yang terkecil (5) . . . . .	25
3.12	Hasil perhitungan jarak terkecil antar kelompok ke objek atau objek ke objek dan mengambil yang terkecil (6) . . . . .	26
3.13	Hasil perhitungan jarak terkecil antar kelompok ke objek atau objek ke objek dan mengambil yang terkecil (7) . . . . .	26
3.14	Hasil perhitungan jarak terkecil antar kelompok ke objek atau objek ke objek dan mengambil yang terkecil (8) . . . . .	26
3.15	Hasil perhitungan manual untuk <i>Agglomerative Hierarchical Clustering</i> . . . . .	26
3.16	Tabel skenario saat pengguna menekan tombol <i>Browse</i> . . . . .	29
3.17	Tabel skenario saat pengguna memilih metode . . . . .	29
3.18	Tabel skenario saat pengguna memasukan jumlah klaster . . . . .	30
3.19	Tabel skenario saat pengguna menekan tombol <i>start clustering</i> . . . . .	30
3.20	Tabel skenario saat pengguna menekan tombol <i>start clustering</i> tetapi file tidak ditemukan . . . . .	30
3.21	Tabel skenario saat pengguna menekan tombol <i>start clustering</i> tetapi tidak menemukan metode untuk melakukan klasterisasi . . . . .	31
3.22	Tabel skenario saat pengguna menekan tombol <i>start clustering</i> tetapi tidak menemukan jumlah klaster yang diinginkan untuk klasterisasi . . . . .	31
3.23	Tabel Skenario saat pengguna menekan tombol <i>exit</i> . . . . .	32
5.1	Tabel <i>intracluster agglomerative hierarchical</i> perhitungan jarak cosine dan dataset bahan tanpa lauk . . . . .	46
5.2	Tabel <i>intracluster</i> terkecil . . . . .	46

5.3	Hasil rata-rata terbesar berdasarkan <i>intracluster distance</i> dengan k=5 pada cluster 3	46
5.4	Daerah-daerah pada k=5	47
5.5	Tabel <i>intracluster K-Means</i> Perhitungan Jarak Cosine dan Dataset Bahan tanpa Lauk	47
5.6	Tabel <i>intracluster</i> terkecil	48
5.7	Hasil rata-rata terbesar berdasarkan <i>intracluster distance</i> dengan k=7 pada klaster 5	48
5.8	Hasil rata-rata terbesar berdasarkan <i>intracluster distance</i> dengan k=7 pada klaster 5 (2)	49
5.9	Daerah-daerah pada k=7	49
5.10	Daerah-daerah pada k=7	50
5.11	Tabel <i>intracluster Agglomerative Hierarchical</i> Perhitungan Jarak <i>Euclidean</i> dan Dataset Bahan tanpa Lauk	50
5.12	Tabel <i>intracluster</i> terkecil	50
5.13	Hasil rata-rata terbesar berdasarkan <i>intracluster distance</i> dengan k=3 pada klaster 2	51
5.14	Daerah-daerah pada k=3	51
5.15	Tabel <i>intracluster K-Means</i> Perhitungan Jarak <i>Euclidean</i> dan Dataset Bahan tanpa Lauk	51
5.16	Tabel <i>intracluster</i> terkecil	52
5.17	Hasil rata-rata terbesar berdasarkan <i>intracluster distance</i> dengan k=5 pada klaster 3	52
5.18	Daerah-daerah pada k=5	53
5.19	Tabel <i>Intracluster Agglomerative Hierarchical</i> Perhitungan Jarak <i>Cosine</i> dan Dataset Bahan dengan Lauk	53
5.20	Tabel <i>intracluster</i> terkecil	53
5.21	Hasil rata-rata terbesar berdasarkan <i>intracluster distance</i> dengan k=6 pada klaster 4	54
5.22	Hasil rata-rata terbesar berdasarkan <i>intracluster distance</i> dengan k=6 pada cluster 4(2)	54
5.23	Daerah-daerah pada k=6	55
5.24	Daerah-daerah pada k=6	55
5.25	Tabel <i>Intracluster K-Means</i> Perhitungan Jarak <i>Cosine</i> dan Dataset Bahan dengan Lauk	56
5.26	Tabel <i>intracluster</i> terkecil	56
5.27	Hasil rata-rata terbesar berdasarkan <i>intracluster distance</i> dengan k=4 pada klaster 2	56
5.28	Daerah-daerah pada k=4	57
5.29	Tabel <i>Intracluster Agglomerative Hierarchical</i> Perhitungan Jarak <i>Euclidean</i> dan Dataset Bahan dengan Lauk	57
5.30	Tabel <i>intracluster</i> terkecil	57
5.31	Hasil rata-rata terbesar berdasarkan <i>intracluster distance</i> dengan k=4 pada klaster 3	58
5.32	Daerah-daerah pada k=4	58
B.1	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan tanpa lauk	75
B.2	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan tanpa lauk (2)	76
B.3	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan tanpa lauk (3)	77
B.4	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan tanpa lauk (4)	78
B.5	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan tanpa lauk (5)	79
B.6	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan tanpa lauk (6)	80
B.7	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan tanpa lauk (7)	81

B.8	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan tanpa lauk	82
B.9	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan tanpa lauk (2)	83
B.10	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan tanpa lauk(3)	84
B.11	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan tanpa lauk(4)	85
B.12	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan tanpa lauk(5)	86
B.13	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan tanpa lauk(6)	87
B.14	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan tanpa lauk(7)	88
B.15	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan dengan lauk	89
B.16	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan dengan lauk(2)	90
B.17	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan dengan lauk(3)	91
B.18	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan dengan lauk(4)	92
B.19	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan dengan lauk(5)	93
B.20	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan dengan lauk(6)	94
B.21	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan dengan lauk(7)	95
B.22	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan dengan lauk	97
B.23	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan dengan lauk(2)	98
B.24	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan dengan lauk(3)	99
B.25	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan dengan lauk(4)	100
B.26	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan dengan lauk(5)	101
B.27	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan dengan lauk(6)	102
B.28	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan dengan lauk(7)	103
B.29	<i>Agglomerative hierarchical</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan dengan lauk(8)	104
B.30	<i>K-Means</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan tanpa lauk	106
B.31	<i>K-Means</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan tanpa lauk(2)	107
B.32	<i>K-Means</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan tanpa lauk(3)	108
B.33	<i>K-Means</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan tanpa lauk(4)	109
B.34	<i>K-Means</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan tanpa lauk(5)	110
B.35	<i>K-Means</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan tanpa lauk(6)	111
B.36	<i>K-Means</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan tanpa lauk(7)	112
B.37	<i>K-Means</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan tanpa lauk	115

B.38	<i>K-Means</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan tanpa lauk (2)	116
B.39	<i>K-Means</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan tanpa lauk (3)	117
B.40	<i>K-Means</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan tanpa lauk (4)	118
B.41	<i>K-Means</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan tanpa lauk (5)	119
B.42	<i>K-Means</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan tanpa lauk (6)	120
B.43	<i>K-Means</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan tanpa lauk (7)	121
B.44	<i>K-Means</i> dengan perhitungan jarak <i>euclidean</i> , fitur bahan tanpa lauk (8)	122
B.45	<i>K-Means</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan dengan lauk	123
B.46	<i>K-Means</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan dengan lauk(2)	124
B.47	<i>K-Means</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan dengan lauk(3)	125
B.48	<i>K-Means</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan dengan lauk(4)	126
B.49	<i>K-Means</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan dengan lauk(5)	127
B.50	<i>K-Means</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan dengan lauk(6)	128
B.51	<i>K-Means</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan dengan lauk(7)	129
B.52	<i>K-Means</i> dengan perhitungan jarak <i>cosine</i> , fitur bahan dengan lauk(8)	130

## DAFTAR KODE PROGRAM

2.1	Menjalankan library BeautifulSoup dan request . . . . .	5
2.2	Cookie dan User-Agent . . . . .	5
2.3	HTML Tag . . . . .	6
2.4	HTML Tag untuk data lebih rinci . . . . .	6
2.5	import CSV . . . . .	6
3.1	Bagian kode program untuk user-agent dan cookie . . . . .	17
3.2	Bagian kode program untuk mengambil link dari setiap daerah di kategori makanan Indonesia . . . . .	17
3.3	Bagian kode program untuk mengambil tag daerah . . . . .	17
3.4	Bagian kode program untuk mengambil tag nama resep . . . . .	17
3.5	Bagian kode program untuk mengambil tag bahan dan kuantitas . . . . .	18
3.6	filtering nama . . . . .	18
3.7	filtering bahan . . . . .	18
A.1	webScraping.py . . . . .	65
A.2	filternama.py . . . . .	66
A.3	filterbahan.py . . . . .	66
A.4	normalisasi.py . . . . .	67
A.5	hirar.py . . . . .	67
A.6	KMeans.py . . . . .	68
A.7	Algorithm.py . . . . .	69
A.8	main.py . . . . .	69

# BAB 1

## PENDAHULUAN

Bab pendahuluan berisi latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi penelitian, dan sistematika pembahasan dari penelitian ini.

### 1.1 Latar Belakang

Sekarang ini sangat mudah untuk mencari dan mendapatkan resep makanan. Resep bisa didapatkan dari banyak tempat, contohnya buku, majalah, maupun dari situs resep makanan di internet. Resep berfungsi untuk membantu orang-orang dalam membuat makanan dengan benar dan sesuai dengan takaran komposisi setiap bahan yang digunakan. Resep berisi bahan-bahan yang digunakan untuk membuat suatu makanan dan tahapan cara memasak dari awal hingga makanan tersebut dapat dikonsumsi. Ada 4 teori dari resep [5] yaitu *realism* mengatakan identitas resep adalah berdasarkan bahan dan tata cara pembuatannya, *constructivism* mengatakan identitas resep bergantung pada siapa yang membuat resep tersebut, *existentialism* mengatakan identitas resep bergantung pada dialog yang relevan dengan individu, dan *naive approach* mengatakan resep adalah produk budaya populer.

Orang-orang dapat dengan mudah menulis berbagai macam resep di internet yang kemudian dimasukkan ke situs-situs yang dikhususkan untuk resep ataupun yang memiliki fitur untuk resep. Saat pengguna melihat sebuah resep yang ada di sebuah laman di internet, ada kemungkinan pengguna tersebut ingin melihat resep yang sejenis atau mirip dengan resep yang dilihat sebelumnya. Untuk mendapatkan resep yang sejenis, salah satu yang dapat dilakukan adalah dengan mengelompokkan resep. Untuk melakukan pengelompokkan guna mencari resep yang berhubungan satu sama lain, diperlukan teknik pengelompokkan yang dinamakan klusterisasi. Klusterisasi merupakan teknik yang mengelompokkan objek-objek dengan karakteristik serupa ke dalam satu kluster dan objek-objek lain yang tidak memiliki karakteristik serupa ke dalam kluster yang berbeda [4]. Ada banyak cara untuk melakukan klusterisasi, beberapa di antaranya adalah *k-means* dan *agglomerative hierarchical clustering*.

*K-means clustering* adalah metode di mana seluruh objek akan dibagi menjadi  $k$  buah kluster [4]. Masing-masing kluster memiliki sebuah *centroid* yang menjadi titik tengah untuk menentukan setiap objek akan berada di dalam kluster mana. *Agglomerative hierarchical clustering* adalah metode klusterisasi yang mengelompokkan data menjadi sebuah hirarki atau pohon [4]. *Agglomerative hierarchical clustering* akan memperlakukan setiap objek menjadi sebuah kluster. Kluster kemudian akan digabungkan dengan kluster lain yang mirip berdasarkan jarak antar dua kluster, penggabungan akan terus dilakukan hingga membentuk satu kluster saja. Kedua teknik klusterisasi ini kemudian akan dibandingkan untuk melihat teknik klusterisasi mana yang cocok untuk mengelompokkan resep. Salah satu cara untuk mengukur performa dari klusterisasi adalah melihat jarak dari *intracluster*. *intracluster* adalah jarak antar objek-objek dalam kluster, semakin kecil jarak *intracluster* maka semakin baik kerapatan dalam kluster yang menandakan bahwa objek-objek dalam cluster memiliki tingkat kemiripan yang baik.



Untuk melakukan klusterisasi atau pengelompokan resep makanan diperlukan data resep-resep makanan terlebih dahulu. Ada banyak cara untuk mendapatkan data resep, salah satunya adalah dengan menggunakan *Web Scraping*. *Web scraping* adalah proses untuk mengekstrak data dari sebuah situs web atau *web data extraction* [1]. Terdapat banyak situs menyediakan data resep makanan, pada penelitian ini situs Cookpad akan digunakan untuk mendapatkan data resep. Dari *website* Cookpad bisa didapatkan nama resep, daerah asal resep, bahan yang digunakan, komposisi dari setiap bahan, dan metode pembuatan makanan. Pada penelitian ini dataset yang digunakan adalah nama resep, daerah asal resep, dan bahan beserta komposisi bahan yang digunakan berdasarkan teori identitas resep yaitu resep bergantung pada bahan yang digunakan berdasarkan teori (*realism*). Setelah data dikumpulkan, data harus disaring dan dibersihkan terlebih dahulu dengan cara menghapus data yang tidak relevan seperti data yang terduplikat, mengisi nilai data yang hilang, dan menghapus atau mengganti data yang tidak tepat. Tujuan dari pembersihan data adalah untuk membuat data lebih baik dan relevan saat diolah. Dari data yang sudah dibersihkan, informasi dari bahan dan komposisi dari bahan yang digunakan akan dijadikan sebagai fitur.

Pada penelitian ini dibuat perangkat lunak untuk mengelompokkan resep-resep agar pengguna dapat melihat resep yang berhubungan satu-sama lain dari Cookpad. Untuk mengelompokkan resep makanan akan digunakan teknik klusterisasi. Klusterisasi akan diimplementasikan menggunakan bahasa pemrograman Python dan hasil dari klusterisasi akan ditampilkan ke dalam bentuk tabel. Selain menggunakan *intracluster* untuk mengukur kesamaan dalam kluster, analisa juga dilakukan dengan melihat isi dari kluster berdasarkan segi cita rasa, jenis masakannya, dan daerah dengan melihat penyebaran dari setiap daerah dan apakah isi dari kluster memiliki ciri kemiripan yang menyebabkan resep-resep makanan bisa berada pada kluster yang sama. Ada penelitian yang meneliti tentang adanya kedekatan pada sifat makanan di beberapa daerah, sehingga jika ada pengguna yang menyukai sebuah makanan di suatu daerah, maka berdasarkan sifat pada makanan bisa diberi rekomendasi makanan di daerah yang sejenis. Pada penelitian ini, setelah pengelompokkan berhasil dilakukan, maka akan dilihat apakah pada hasil klusterisasi terdapat daerah-daerah yang selalu berada dalam satu kelompok. Jika ada maka dapat dilakukan rekomendasi berdasarkan daerah asal makanan tersebut, tetapi jika tidak maka rekomendasi tidak bisa dilakukan.

## 1.2 Rumusan Masalah

Adapun rumusan masalah dari deskripsi yang telah dipaparkan adalah:

1. Bagaimana cara membuat vektor fitur yang akan digunakan untuk proses klusterisasi?
2. Bagaimana membangun perangkat lunak untuk melakukan klusterisasi resep makanan Indonesia berdasarkan bahan yang digunakan dan komposisi bahan makanan?
3. Bagaimana melakukan analisis untuk hasil kluster yang diperoleh?

## 1.3 Tujuan

Adapun tujuan penelitian dari rumusan masalah yang telah dipaparkan adalah:

1. Membuat vektor fitur untuk digunakan dalam proses klusterisasi.
2. Membangun perangkat lunak untuk melakukan klusterisasi resep makanan Indonesia.
3. Menganalisa hasil dari klusterisasi.

## 1.4 Batasan Masalah

Berikut adalah batasan masalah pada penelitian ini :

Penelitian ini menggunakan dataset yang berisi nama dari resep, daerah asal resep, bahan yang digunakan, dan kuantitas dari setiap bahan yang digunakan pada resep.

## 1.5 Metodologi

1. Melakukan studi literatur tentang resep.
2. Melakukan studi literatur tentang metode *clustering k-means* dan *agglomerative hierarchical*.
3. Mengumpulkan data-data resep makanan Indonesia.
4. Melakukan pra-pemrosesan untuk dataset.
5. Menganalisis masalah perangkat lunak yang akan dibuat.
6. Membuat desain atau rancangan perangkat lunak yang akan dibuat.
7. Mengimplementasikan perangkat lunak yang akan dibuat.
8. Pengujian hasil dari proses klasterisasi.
9. Menarik kesimpulan berdasarkan hasil keluaran perangkat lunak.

## 1.6 Sistematika Pembahasan

Berikut adalah sistematika pada setiap bab dalam penelitian ini :

1. Bab 1 atau Pendahuluan, membahas mengenai latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, metodologi penelitian, dan sistematika pembahasan.
2. Bab 2 atau Landasan Teori, membahas teori-teori yang akan digunakan sebagai landasan dari penelitian ini yaitu penjelasan tentang *web scraping*, teknik *data mining*, kemiripan antar vektor, normalisasi data, klasterisasi, *intracluster*, dan resep sebagai landasan penelitian.
3. Bab 3 atau Analisis Perangkat Lunak, membahas tentang analisis masalah, analisis dataset, studi kasus sederhana, dan gambaran umum perangkat lunak.
4. Bab 4 atau Perancangan Perangkat Lunak, membahas tentang perancangan kelas yang ada pada perangkat lunak dan perancangan antarmuka untuk perangkat lunak.
5. Bab 5 atau Implementasi dan Pengujian, membahas tentang implementasi dan pengujian dari perangkat lunak klasterisasi resep makanan Indonesia menggunakan metode klasterisasi *k-means* dan *agglomerative hierarchical*.
6. Bab 6 atau Kesimpulan dan Saran, membahas tentang kesimpulan dari penelitian yang didapatkan dan saran untuk pengembangan penelitian.