

# SISTEM INFORMASI GEOGRAFIS BERBASIS WEB UNTUK PRODUKSI KEDELAI DI INDONESIA MENGGUNAKAN K-MEANS CLUSTERING

## WEB-BASED GEOGRAPHIC INFORMATION SYSTEM FOR SOYBEAN PRODUCTION IN INDONESIA USING K-MEANS CLUSTERING

Resti Hidayah <sup>1\*</sup>, Imas Sukaesih Sitanggang <sup>2</sup>

Program Studi S1 Ilmu Komputer, Departemen Ilmu Komputer, FMIPA, Institut Pertanian Bogor,  
Bogor <sup>1\*</sup>

restihidayah5@gmail.com

Kampus IPB Dramaga, Jl. Meranti Wing 20 Level V, Bogor, Indonesia 16680

Departemen Ilmu Komputer, FMIPA, Institut Pertanian Bogor, Bogor <sup>2</sup>

### ABSTRACT

Indonesia has large and potential area for soybean cultivation. Information concerning areas that are potential for soybean production is not provided in an interactive information system that can be easily accessed by public. Nowadays, information about soybean production in all districts in Indonesia is displayed in the tabular form in the website of Ministry of Agriculture. For that, in this study, a web-based geographic information system (GIS) was developed to manage and visualize soybean production in all districts in Indonesia. K-means clustering was performed to group data on production, productivity and harvested areas of soybean. The best clustering has the number of cluster 4 and results sum square error (SSE) 2.07, 1.53, and 0.0037 for data on production, productivity, and harvested areas of soybean, respectively. The GIS was developed using OpenGeo Suite 3.0 as the geospatial tool for managing and providing information about soybean production in Indonesia in forms of maps, tables, and graphs based on the results of k-means clustering.

*Keywords: geographic information system, clustering, k-means, soybean production*

### ABSTRAK

Indonesia memiliki lahan dengan potensi tinggi dan cukup luas untuk ditanami kacang kedelai. Namun informasi mengenai wilayah-wilayah yang berpotensi untuk produksi kedelai belum disajikan dalam sistem informasi yang interaktif dan mudah diakses dengan mudah oleh masyarakat. Saat ini informasi produksi kedelai di setiap kabupaten di Indonesia disajikan dalam bentuk tabel pada situs Departemen Pertanian. Oleh karena itu, dalam penelitian ini sistem informasi geografis (SIG) berbasis web dibangun untuk memudahkan pengguna memperoleh informasi produksi kedelai di seluruh kabupaten di Indonesia. Teknik *clustering* k-means diterapkan untuk mengelompokan data produksi, produktivitas, dan luas panen kedelai untuk selanjutnya disajikan dalam SIG. *Clustering* terbaik dengan jumlah cluster 4 menghasilkan *sum square error* (SSE) sebesar 2.07, 1.53, dan 0.0037 berturut-turut untuk data luas panen, produksi, dan produktivitas. SIG dibangun dengan menggunakan OpenGeo Suite 3.0 sebagai perangkat lunak geospasial untuk mengelola dan menyediakan informasi mengenai produksi kedelai di Indonesia dalam bentuk peta, tabel, dan grafik berdasarkan hasil k-means *clustering*.

*Katakunci: sistem informasi geografis, clustering, k-means, produksi kedelai*

---

## PENDAHULUAN

Kacang - kacangan banyak dibutuhkan sebagai bahan baku industri pangan, salah satunya yaitu kacang kedelai. Kacang kedelai bisa digunakan sebagai bahan baku pembuatan kecap, susu kedelai, tahu, dan tempe. Selain itu kacang kedelai juga memiliki nilai gizi yang cukup tinggi yang mengandung protein, karbohidrat, lemak, kalsium, dan vitamin B kompleks.

Jumlah produksi kacang kedelai di Indonesia dapat dikatakan belum mampu memenuhi kebutuhan kedelai dalam negeri. Berdasarkan data dari Departemen Pertanian Indonesia pada tahun 2012, Indonesia hanya melakukan ekspor kacang kedelai sebesar 67,899,346 kg, sedangkan impor kacang kedelai mencapai 4,257,526,480 kg. Hal ini terjadi karena tingginya kebutuhan akan kacang kedelai di dalam negeri setiap tahunnya. Untuk tempe dan tahu saja, kebutuhan kedelai sangat tinggi, ditambah lagi kebutuhan kedelai untuk bahan baku makanan ternak. Karena tingginya tingkat kebutuhan akan kacang kedelai maka negara kita masih tergantung akan impor kedelai. Bahkan karena kebutuhan yang terus meningkat maka nilai impor kacang kedelai setiap tahunnya juga terus meningkat [1].

Bila melihat potensi wilayah serta lahan yang dimiliki Indonesia maka tidak seharusnya Indonesia selalu mengimpor kedelai untuk memenuhi kebutuhan pangan dalam negeri. Begitu banyak provinsi di Indonesia yang memiliki lahan berpotensi tinggi dan cukup luas untuk ditanami kacang kedelai, diantaranya yaitu Jawa Timur, Jawa Tengah, Sumatera Barat, Papua Barat, Jawa Barat, dan Sulawesi Selatan, sedangkan lahan berpotensi sedang untuk ditanami kedelai ada di Lampung, NAD, Banten, Nusa Tenggara Barat, dan Sulawesi Tenggara [2]. Namun banyaknya wilayah Indonesia yang berpotensi ini tidak didukung oleh adanya informasi yang dapat dilihat oleh masyarakat Indonesia. Oleh karena itu diperlukan adanya suatu sistem informasi geografis (SIG) untuk menampilkan informasi mengenai produksi kedelai wilayah – wilayah di Indonesia. SIG merupakan salah satu sistem yang memungkinkan kita untuk mengelola data spasial menjadi sebuah informasi yang eksplisit, dan digunakan untuk membuat sebuah keputusan [3].

Pada penelitian ini, teknik *clustering* k-means diterapkan untuk mengelompokkan data produksi, produktivitas, dan luas panen kedelai untuk selanjutnya disajikan dalam SIG. K-means merupakan metode pengelompokkan data nonhierarki yang mempartisi data berdasarkan nilai pusat dari *cluster*. Metode ini mempartisi data kedalam dua kelompok atau lebih, sehingga data yang berkarakteristik sama dimasukkan kedalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan kedalam kelompok yang lain [4]. SIG berbasis web untuk produksi kedelai dibangun menggunakan OpenGeo Suite 3.0 sebagai salah satu perangkat lunak geospasial untuk mengelola dan menyediakan informasi mengenai produksi kedelai di Indonesia dalam bentuk peta, tabel, dan grafik.

## METODE PENELITIAN

### Data, Alat, dan Area Studi

Pada penelitian ini, data spasial yang digunakan adalah data vektor untuk peta administrasi kabupaten dari 33 provinsi di Indonesia dan data atributnya adalah nama kabupaten, luas panen, jumlah produksi, produktivitas, dan tahun produksi kacang kedelai di 33 provinsi di Indonesia dari tahun 2000-2011 yang diperoleh dari Departemen

Pertanian Republik Indonesia (<http://aplikasi.deptan.go.id>). Data awal yang diperoleh masih berupa file dengan format .xls sehingga harus dilakukan praproses data terlebih dahulu sebelum data diolah untuk membangun SIG. Data yang telah diproses akan dikelompokkan menggunakan perangkat lunak Weka 3.6.9 dan selanjutnya diolah menggunakan OpenGeo Suite 3.0 sebagai salah satu perangkat lunak geospasial untuk mengelola dan menyediakan informasi mengenai produksi kedelai di Indonesia dalam bentuk peta, tabel, dan grafik.

### **Tahapan Penelitian**

#### Praproses

Pada penelitian ini dilakukan 2 tahapan praproses data, yaitu: pembersihan data dan pemilihan data. Pembersihan data dilakukan untuk menghapus data yang tidak konsisten dan *outlier* yang akan mengganggu pada tahap *data mining* atau mengisi nilai *missing value*. Pemilihan data dilakukan untuk memilih variabel – variabel yang akan digunakan dalam tahap *clustering*.

#### Clustering

*Clustering* adalah proses mengelompokkan data (objek) yang didasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan di antaranya, sehingga data yang berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain. Pada penelitian ini, teknik *clustering* dilakukan untuk mengelompokkan data produksi kedelai berdasarkan nilai luas panen, produksi, dan produktivitas. Teknik *clustering* yang diterapkan adalah algoritme k-means. Berikut adalah algoritme dari k-means [4]:

#### *Input:*

- k: jumlah *cluster*
- D: *Dataset* yang berisi n objek

#### *Output:*

Himpunan yang terdiri dari k *cluster*

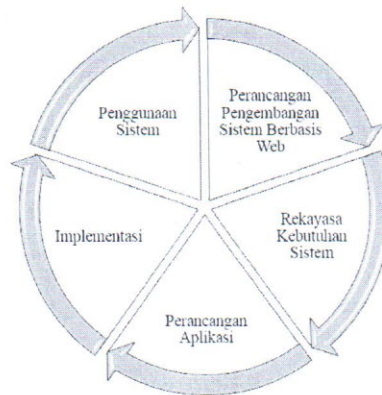
#### *Method:*

- (1) pilih k objek secara acak dari D sebagai pusat *cluster* awal
- (2) ulangi
- (3) tempatkan setiap objek ke *cluster* yang paling dekat berdasarkan nilai rata-rata dari objek dalam *cluster*
- (4) perbaharui nilai rata – rata *cluster* dengan menghitung nilai rata-rata dari objek untuk setiap *cluster*
- (5) hingga tidak terjadi perubahan pada *cluster*

#### Pembuatan Sistem Informasi Geografis

Pembuatan Sistem Informasi Geografis dalam penelitian ini dilakukan melalui tahapan dalam metode Web-based Development Life Cycle (WDLC) yang disajikan pada Gambar 1. Metode ini terdiri dari 5 tahapan, yaitu: perencanaan pengembangan sistem berbasis web, rekayasa kebutuhan sistem, perancangan aplikasi, implementasi, dan penggunaan sistem [5]

---



Gambar 14 Tahapan penelitian [5]

## HASIL DAN PEMBAHASAN

### Praproses

Data awal yang diperoleh adalah *file* dengan format *.xls* yang berisi nilai jumlah produksi, luas panen, dan produktivitas kedelai pada setiap kabupaten di Indonesia dari tahun 2000 – 2011. Pada *file* tersebut masih terdapat beberapa *missing value* pada nilai produktivitas dan terdapat beberapa kesalahan perhitungan nilai produktivitas pada data. Oleh karena itu diperlukan tahapan pembersihan data untuk mengatasi hal tersebut. Untuk memperbaiki dan mengisi nilai produktivitas yang kosong, dilakukan perhitungan berdasarkan jumlah produksi dan luas panen sesuai persamaan (1) berikut.

$$\text{Produktivitas} = \frac{\text{Jumlah Produksi (Kuintal)}}{\text{Luas Panen (Ha)}} \quad (1)$$

Selanjutnya, dilakukan tahapan pemilihan data untuk memilih variabel – variabel yang diperlukan. Pada tahapan ini, variabel – variabel yang dipilih yaitu variabel tahun, nama kabupaten, jumlah produksi, luas panen, dan produktivitas.

### Clustering

*Clustering* dilakukan untuk mengelompokkan data nilai produksi, luas panen, dan produktivitas pada setiap kabupaten. *Clustering* k-means dilakukan dengan menggunakan perangkat lunak Weka 3.6.9 untuk beberapa nilai k, yaitu 2, 3, 4, dan 5. Hasil clustering pada data luas panen, produksi dan produktivitas berturut-turut dapat dilihat pada Tabel 1, 2, dan 3.

Tabel 1 Hasil *clustering* luas panen

Jumlah K	SSE (Sum Square Error)	Cluster	Jumlah Anggota	Luas Panen		
				Maks	Min	Rata - Rata
2	6.407	C1	2,909	11,158	1	865
		C2	138	49,056	11,432	21,667

3	3.66	C1	266	14,479	3,263	6,076.95
		C2	116	49,056	14,912	23,347.94
		C3	2,665	3,233	1	443.24
4	2.07	C1	291	11,783	2,853	5,222.70
		C2	98	24,539	11,994	18,615.69
		C3	2,623	2,804	1	401.86
		C4	35	49,056	25,165	31,652.97
5	1.29	C1	63	15,233	7,518	10,887.49
		C2	91	26,939	15,873	20,790.26
		C3	331	7,307	2,142	3,943.55
		C4	22	49,056	28,068	35,052.27
		C5	2,540	2,132	1	335.34

Tabel 2 Hasil *clustering* produksi

Jumlah K	SSE ( <i>Sum Square Error</i> )	Cluster	Jumlah Anggota	Produksi		
				Maks	Min	Rata - Rata
2	5.57	C1	130	2,983	1	858.85
		C2	2,921	14,588	1	1,171.33
3	3.37	C1	95	76,434	21,182	32,354.85
		C2	226	20,916	5,285	9,898.42
		C3	2,730	5,251	1	667.49
4	1.53	C1	18	76,434	40,985	53,005.38
		C2	274	16,113	3,988	7,416.21
		C3	2,652	3,957	1	553.25
		C4	107	38,514	16,451	25,043.31
5	1.02	C1	18	76,434	40,985	53,005.38
		C2	350	9,571	2,714	5,004.78
		C3	2,517	2,699	1	407.42
		C4	88	20,655	9,793	14,332.06
		C5	78	38,514	20,916	27,442.69

Tabel 3 Hasil *clustering* produktivitas

Jumlah K	SSE ( <i>Sum Square Error</i> )	Cluster	Jumlah Anggota	Produktivitas		
				Maks	Min	Rata - Rata
2	0.078	C1	1	11,170	11,170	11,170
		C2	2,983	1,434.9	0.09	17.91
3	0.015	C1	1	11,170	11,170	11,170
		C2	6	1,434.9	970.83	1,163.02
		C3	2,977	471.65	0.09	15.60
4	0.0037	C1	1	11,170	11,170	11,170

		C2	6	1,434.9	970.83	1,163.02
		C3	77	471.65	80.93	147.94
		C4	2,900	79.02	0.09	12.09
5	0.0036	C1	1	11,170	11,170	11,170
		C2	77	471.65	80.93	147.94
		C3	2,797	79.02	7.78	12.42
		C4	103	7.69	0.09	3.07
		C5	6	1,434.9	970.83	1,163.02

Pada Tabel 1, 2, dan 3 bisa dilihat jumlah k, nilai SSE, jumlah anggota, nilai maksimum, minimum, dan rata-rata dari setiap hasil *clustering*. Berdasarkan hasil *clustering* tersebut diperoleh hasil *cluster* terbaik dengan menggunakan k=4 karena dengan jumlah *cluster* yang tidak terlalu banyak telah menghasilkan nilai SSE yang cukup kecil. Dengan menggunakan nilai maksimum, minimum, dan rata-rata dengan k=4 pada setiap hasil *clustering* ditentukan rentang nilai untuk masing-masing *cluster* pada nilai luas panen, produksi, dan produktivitas, seperti dapat dilihat pada Tabel 4.

Tabel 4 Rentang nilai hasil *clustering*

Cluster	Rentang Nilai		
	Luas Panen (Ha)	Produksi (Ton)	Produktivitas (Ku /Ha)
Rendah	<2850	<3980	<80
Sedang	>=2850 dan <12000	>=3980 dan <16400	>=80 dan <970
Tinggi	>=12000 dan <25000	>=16400 dan <40900	>=970 dan <11160
Sangat Tinggi	>=25000	>=40900	>=11160

Tabel 4 merupakan rentang nilai yang diperoleh untuk masing-masing *cluster* pada nilai luas panen, produksi, dan produktivitas. Dengan menggunakan rentang nilai ini, maka dilakukan proses transformasi data untuk mengubah nilai numerik pada setiap nilai pada luas panen, produksi, dan produktivitas menjadi nilai kategori rendah, sedang, tinggi, dan sangat tinggi. Sementara untuk nilai yang bernilai 0 dan na masing-masing diberi nilai Null dan Na. Tabel 5 adalah contoh data kategori produksi kedelai tahun 2000 berdasarkan hasil *clustering*.

Tabel 5 Contoh data kategori produksi kedelai tahun 2000 berdasarkan hasil *clustering*

Nama kab	Luas Panen (Ha)	Kategori Luas Panen	Produksi (Ton)	Kategori Produksi	Produktivitas (Ku/Ha)	Kategori Produktivitas
Simeulue	Null	Null	Null	Null	Null	Null
Aceh Singkil	Null	Null	Null	Null	Null	Null
Aceh Selatan	363	Rendah	Null	Null	Null	Null
Aceh Tenggara	119	Rendah	155	Rendah	13.03	Rendah
Aceh Timur	6,148	Sedang	7,277	Sedang	11.84	Rendah
Aceh Tengah	254	Rendah	283	Rendah	11.14	Rendah
Aceh Barat	3,385	Sedang	4,017	Sedang	11.87	Rendah
Aceh Besar	288	Rendah	384	Rendah	13.33	Rendah
Pidie	Na	Na	Na	Na	Na	Na

Bireuen	Null	Null	Null	Null	Null	Null
Aceh Utara	48,212	Sangat Tinggi	58,814	Sangat Tinggi	12.2	Rendah
Aceh Barat Daya	Null	Null	Null	Null	Null	Null

## **Pembuatan Aplikasi**

### Perancangan Pengembangan Sistem Berbasis Web

Pada tahap ini dilakukan perencanaan pembuatan SIG untuk produksi kedelai yang mampu memenuhi tujuan dari penelitian ini, yaitu SIG yang mampu mengelola data spasial dan data tekstual, serta mampu menampilkan informasi peta, grafik, dan tabel dari hasil produksi kedelai di 33 provinsi di Indonesia.

### Rekayasa Kebutuhan Sistem

Pada tahap ini dilakukan pengumpulan serta menganalisis semua kebutuhan SIG untuk data produksi kedelai. Kebutuhan fungsional yang dibutuhkan untuk SIG ini adalah fungsi menampilkan peta, tabel, dan grafik.

### Perancangan Aplikasi

Setelah semua kebutuhan fungsional dikumpulkan, dilakukan tahap perancangan aplikasi. Perancangan aplikasi terdiri dari 2 bagian, yaitu perancangan kebutuhan data produksi kedelai dan perancangan kebutuhan fungsional. Perancangan kebutuhan data produksi kedelai menghasilkan data produksi kedelai yang telah dikelompokkan menggunakan algoritme k-means, sedangkan perancangan kebutuhan fungsional menghasilkan gambaran umum mengenai fungsi menampilkan peta, tabel, dan grafik.

### Implementasi

Tahap ini dilakukan untuk merealisasikan rancangan SIG berbasis web untuk produksi kedelai yang telah dibuat dengan menggunakan OpenGeo Suite 3.0. Pembuatan basis data spasial produksi kedelai dilakukan dengan menggunakan PostgreSQL. Pembuatan *layer* dan *style* peta dilakukan menggunakan OpenLayer. Pengaturan *style* dilakukan untuk memberikan pewarnaan pada setiap kabupaten berdasarkan kategori produksi, luas panen, dan produktivitas yang diperoleh dari hasil *clustering* menggunakan algoritme k-means. Warna yang digunakan adalah hitam, abu – abu, hijau, kuning, jingga, dan merah dimana masing – masing warna mewakili kategori na, null, rendah, sedang, tinggi, dan sangat tinggi. Serta untuk membangun dan mengimplementasikan kode program dilakukan dengan menggunakan *framework CodeIgniter*. Gambar 2 menunjukkan tampilan peta dan tabel produksi kedelai untuk kategori luas panen pada tahun 2003 dari semua provinsi di Indonesia yang dihasilkan berdasarkan hasil *clustering* menggunakan algoritme k-means.



Gambar 2 Tampilan SIG produksi kedelai

Pada Gambar 2, tampilan peta, warna hijau, kuning, jingga, dan merah menunjukkan hasil *clustering* luas panen yang bernilai rendah, sedang, tinggi, dan sangat tinggi, sedangkan warna hitam dan abu – abu menunjukkan nilai luas panen yang bernilai na dan null. Sedangkan tampilan tabel menjelaskan nilai luas panen dan kategori luas panen dari setiap kabupaten di seluruh Indonesia.

#### Penggunaan Sistem

Pada tahap ini, SIG untuk produksi kedelai yang telah dibuat dan diuji dipublikasikan, serta dilakukan pemeliharaan untuk memastikan isi dari SIG tetap sesuai dengan kebutuhan dan spesifikasi sistem. Proses pemeliharaan merupakan proses yang berkelanjutan dimana sistem secara terus menerus diperbarui dan disesuaikan dengan perubahan.

#### KESIMPULAN

Sistem informasi geografis berbasis web untuk produksi kedelai di Indonesia berdasarkan hasil *clustering* menggunakan algoritme k-means telah berhasil dibangun



menggunakan perangkat lunak OpenGeo Suite 3.0 dan *framework CodeIgniter*. Hasil *clustering* terbaik diperoleh dengan jumlah *cluster* 4 dan menghasilkan *sum square error* (SSE) sebesar 2.07, 1.53, dan 0.0037 berturut-turut untuk data luas panen, produksi, dan produktivitas. Berdasarkan hasil *clustering*, data produksi kedelai dibagi ke dalam 4 kategori yang selanjutnya ditampilkan dalam peta. Selain dalam bentuk peta, sistem ini juga mampu menampilkan hasil produksi kedelai dalam bentuk tabel dan grafik.

## PUSTAKA

- [1] Nazaruddin. 1993. *Komoditi Ekspor Pertanian*. Jakarta: PT Penebar Swadaya.
  - [2] Balai Besar Penelitian dan Pengembangan Sumberdaya Lahan Pertanian. 2008. *Potensi dan Ketersediaan Lahan untuk Pengembangan Kedelai di Indonesia*. *Warta Penelitian dan Pengembangan Pertanian* Vol. 30, No. 1: 3-4.
  - [3] DeMers. 1997. *Fundamentals of Geographic Information Systems*. USA: Hamilton Printing.
  - [4] Han J, Kamber M, and Pei J. 2012. *Data Mining: Concept and Techniques*. Massachusetts: Morgan Kaufmann.
  - [5] Abdul-Aziz A, Koronios A, Gao J, and Sulong MS. 2012. A Methodology for the Development of Web-based Information System: Web Development Team Perspective. *Proceedings of the Eighteenth Americas Conference on Information System*: 1-9.
-