



Hak Cipta Dilindungi Undang-undang  
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah  
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.  
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

## **VISUALISASI HASIL *CLUSTERING DATA BERBASIS MEDIA SOSIAL UNTUK KASUS PERTANIAN DI INDONESIA***

**MUHAMMAD NUR HUSAIN**



**DEPARTEMEN ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2015**

# IPB University

*@Hak cipta milik IPB University*



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah  
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



## **PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA**

Dengan ini saya menyatakan bahwa skripsi berjudul Visualisasi Hasil *Clustering* Data Berbasis Media Sosial untuk Kasus Pertanian di Indonesia adalah benar karya saya dengan arahan dari komisi pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Desember 2015

*Muhammad Nur Husain*  
NIM G64110060

Hak Cipta Dilindungi Undang-undang  
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah

b. Pengutipan titik merujukan kepentingan yang wajar IPB University.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



## ABSTRAK

MUHAMMAD NUR HUSAIN. Visualisasi Hasil *Clustering* Data Berbasis Media Sosial untuk Kasus Pertanian di Indonesia. Dibimbing oleh IMAS SUKAESIH SITANGGANG.

Pertanian merupakan sektor yang signifikan dalam peningkatan kesejahteraan sebagian masyarakat Indonesia. Sektor ini memiliki cakupan yang luas sehingga dibutuhkan sebuah sistem yang mampu mengelola, menyimpan, dan menampilkan informasi beraserensi geografis. Penelitian ini bertujuan membangun sebuah aplikasi web untuk memvisualisasikan hasil *clustering* data media sosial yang telah dilakukan dalam penelitian-penelitian sebelumnya. Aplikasi ini dibangun menggunakan *framework* Shiny dan bahasa pemrograman R. Visualisasi yang dihasilkan berupa peta *cluster*, dendrogram, dan *wordcloud*. Data teks Twitter yang digunakan berjumlah 51 *tweet* yang terbagi ke dalam 11 *cluster* dengan 7 kata kunci yang merepresentasikan sektor pertanian. Visualisasi data teks menampilkan kata kunci yang paling dominan muncul pada data teks Twitter yang digunakan. Adapun peta *cluster* menampilkan hasil data *clustering* dengan warna titik yang berbeda pada setiap *cluster*.

Kata kunci: *clustering*, Shiny, Twitter, *wordcloud*

## ABSTRACT

MUHAMMAD NUR HUSAIN. Data Clustering Result Visualization Based on Social Media for Agriculture Cases in Indonesia. Supervised by IMAS SUKAESIH SITANGGANG.

Agriculture is an important sector for welfare improvement in Indonesia. This sector has a huge scope therefore, a system which is capable of managing, storing, and displaying geographical information is needed. This research was aimed to build a web application to visualize clustering results from social media that have been done in previous researches. This application uses Shiny framework and R programming language. The results are cluster map, dendrogram, and wordcloud. The Twitter text data used consisted of 51 tweets divided into 11 clusters with 7 keywords that represent agricultural sectors. The visualization displayed the most dominant keywords that appear in the text data used in Twitter. The cluster map displayed the results of data clustering with different color points in each cluster.

Keywords: *clustering*, Shiny, Twitter, *wordcloud*



**VISUALISASI HASIL CLUSTERING DATA BERBASIS  
MEDIA SOSIAL UNTUK KASUS PERTANIAN DI  
INDONESIA**

**MUHAMMAD NUR HUSAIN**

Skripsi  
sebagai salah satu syarat untuk memperoleh gelar  
Sarjana Komputer  
pada  
Departemen Ilmu Komputer

**DEPARTEMEN ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2015**



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

**Penguji:**

- 1 Muhammad Abrar Istiadi, SKomp MKom
- 2 Husnul Khotimah, SKomp MKom



Judul Skripsi: Visualisasi Hasil *Clustering* Data Berbasis Media Sosial untuk Kasus Pertanian di Indonesia

Nama : Muhammad Nur Husain  
NIM : G64110060

@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber ;
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Disetujui oleh

Dr Imas Sukaesih Sitanggang, SSi MKom  
Pembimbing

Diketahui oleh



Dr Ir Agus Buono, MSi MKom  
Ketua Departemen

Tanggal Lulus: 22 DEC 2015

**IPB University**



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
- b. Pengutipan titik merugikan kepentingan yang wajar IPB University.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

## **PRAKATA**

Puji dan syukur penulis panjatkan kepada Allah *subhanahu wa ta'ala* atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Januari 2015 ini ialah visualisasi *clustering* data, dengan judul Visualisasi Hasil *Clustering* Data Berbasis Media Sosial untuk Kasus Pertanian di Indonesia.

Selama proses penulisan skripsi ini, penulis banyak mengalami kendala dan masalah, namun berkat bantuan, arahan, serta dukungan dari berbagai pihak, segala kendala dapat terselesaikan. Untuk itu, penulis ucapan terima kasih kepada kedua orang tua dan kedua saudara penulis atas doa, dukungan, dan kasih sayangnya. Terima kasih juga penulis ucapan kepada Ibu Dr Imas Sukaesih Sitanggang, SSi MKom dan Bapak Hari Agung Adrianto, SKom MSi selaku pembimbing, yang senantiasa sabar, tekun, dan ikhlas meluangkan waktunya untuk memberikan bimbingan dan motivasi selama menyusun skripsi.

Penulis juga menyampaikan terima kasih kepada:

- 1 Bapak Muhammad Abrar Istiadi, SKomp MKom dan Ibu Husnul Khotimah, SKomp MKom selaku pengaji.
- 2 Bapak Dr Ir Agus Buono, MSi MKom selaku Ketua Program Studi Ilmu Komputer IPB.
- 3 Seluruh dosen dan staf pegawai tata usaha Departemen Ilmu Komputer IPB.
- 4 Seluruh teman-teman satu bimbingan, yaitu Firdaus, Ryan, Ihda, dan Lalitya.
- 5 M Reza Alfarabi, Ikhsan Nugraha Maulana, Fadhlulrahman Azis, Rizki Adi N, Miftah Farid, Garyndo Arysadewo, M Ega Sanjaya, Randolph Wibowo untuk suka dan duka yang telah dilalui bersama semenjak semester 3.
- 6 Seluruh teman-teman Program S1 Ilmu Komputer angkatan 48.
- 7 Semua pihak yang telah membantu, secara langsung maupun tidak langsung dalam pelaksanaan tugas akhir ini.

Semoga segala bantuan, bimbingan, motivasi, dan dukungan yang telah diberikan kepada penulis senantiasa dibalas oleh Allah *subhanahu wa ta'ala*. Semoga karya ilmiah ini bermanfaat dan menambah wawasan bagi pembaca.

Bogor, Desember 2015

*Muhammad Nur Husain*



<b>DAFTAR TABEL</b>	vi
<b>DAFTAR GAMBAR</b>	vi
<b>DAFTAR LAMPIRAN</b>	vi
<b>PENDAHULUAN</b>	1
Latar Belakang	1
Perumusan Masalah	2
Tujuan Penelitian	2
Manfaat Penelitian	2
Ruang Lingkup Penelitian	2
<b>METODE</b>	2
Data Penelitian	2
Tahapan Penelitian	3
Lingkungan Pengembangan	5
<b>HASIL DAN PEMBAHASAN</b>	6
Data <i>Tweet</i>	6
Praproses Data <i>Tweet</i>	6
Pembuatan Modul <i>Clustering</i>	8
Cluster Data <i>Tweet</i>	8
Pembuatan Modul Visualisasi Hasil <i>Cluster</i>	10
Evaluasi Peta Hasil <i>Clustering</i>	10
Pembuatan Modul <i>Wordcloud</i>	13
Evaluasi Modul <i>Wordcloud</i>	13
Rencana Pengembangan Sistem	15
Kelemahan Aplikasi	16
<b>SIMPULAN DAN SARAN</b>	16
Simpulan	16
Saran	17
<b>DAFTAR PUSTAKA</b>	17
<b>LAMPIRAN</b>	18

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah

b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



## DAFTAR CIPRA MINI DANU MUSAWI

1	Atribut data <i>tweet</i>	3
2	Hasil praproses data <i>clustering</i>	7
	Contoh DF pada data <i>tweet</i>	14
<b>DAFTAR GAMBAR</b>		
3	Gambaran umum sistem informasi geografis berbasis media sosial	3
4	Tahap penelitian	4
7	Contoh data <i>tweet</i> yang digunakan dalam visualisasi	7
8	Potongan program server (kiri) dan antarmuka (kanan)	8
9	Dendrogram tipe <i>complete linkage</i>	9
9	Dendrogram tipe <i>single linkage</i>	9
10	Dendrogram tipe <i>average linkage</i>	10
11	Hasil visualisasi pada peta	11
12	Tampilan <i>pop up</i> yang berisi informasi dari <i>tweet</i>	12
14	Hasil visualisasi modul <i>wordcloud</i>	14

## DAFTAR LAMPIRAN

1	Contoh data <i>tweet</i> tentang pertanian di Indonesia	18
2	Data hasil frekuensi dokumen	20
3	Kode program untuk menampilkan plot berdasarkan tanggal	21
4	Plot hasil <i>clustering</i> data pertanian secara detail lokasi	22

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah  
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



## PENDAHULUAN

### Latar Belakang

Perkembangan dan kemajuan teknologi internet yang ada seperti sekarang sudah semakin pesat tetapi hanya digunakan untuk sekedar *update* status atau juga saling menimpali komentar atau foto yang diunggah ke Twitter (Kemenkominfo 2013). Di Twitter terdapat banyak informasi yang diperoleh, salah satunya di bidang pertanian. Pertanian merupakan sektor yang signifikan untuk meningkatkan kesejahteraan sebagian masyarakat Indonesia. Hal itu merupakan sumber mata pencaharian dari sekitar 60% rakyat Indonesia yang kemudian menjadi sektor riil yang memiliki peran sangat nyata dalam membantu penghasilan devisa negara. Oleh karena itu, dibuatlah sistem informasi yang memanfaatkan Twitter sebagai sumber datanya untuk kasus pertanian di Indonesia.

Susanto (2014) telah melakukan visualisasi data teks Twitter berbasis Bahasa Indonesia dengan menggunakan teknik *clustering* untuk menganalisis sentimen *tweet* mengenai Pemilu 2014. Penelitian ini bertujuan memberikan kemudahan bagi pengguna melihat kecenderungan data berkorelasi positif atau negatif untuk mempermudah tugas-tugas organisasi.

Sugianto (2010) telah melakukan penelitian menggunakan sistem informasi geografis untuk pemetaan dan analisis daerah pertanian di Kabupaten Ponorogo. Sistem ini dibangun menggunakan aplikasi ArcView, MapServer, PostgresSQL, dan QuantumGIS yang diharapkan dapat memberikan informasi secara detail tentang hasil-hasil pertanian yang ada di Kabupaten Ponorogo.

Pada penelitian ini akan dibangun aplikasi web visualisasi berbasis media sosial untuk kasus pertanian. Visualisasi ini memberikan informasi berupa lokasi di mana *tweet* tersebut diakses dan menampilkan isi dari *tweet* tersebut yang berhubungan dengan pertanian serta memvisualisasikannya ke dalam bentuk dendrogram dan *wordcloud*. Data diakuisisi oleh Pratama (2015) dengan menggunakan *application programming interface* (API) dari Twitter berdasarkan empat kategori yaitu (1) *timeline* jika ingin mencari *tweet* berdasarkan *username*, (2) jumlah jika ingin mencari *tweet* global berdasarkan kata kunci sesuai jumlah yang dibutuhkan, (3) waktu jika ingin mencari *tweet* berdasarkan tanggal, dan (4) tempat jika ingin mencari *tweet* berdasarkan lokasi. Setelah data diakuisisi, langkah selanjutnya adalah praproses data. Hasil dari praproses data tersebut diubah oleh Husnayain (2015) kemudian diubah ke dalam bentuk yang sesuai untuk penyimpanan data di MongoDB dan digunakan untuk analisis *cluster* oleh Ahsana (2015). Data yang dipraproses yaitu data teks Twitter yang terbagi menjadi dua jenis, yaitu data *tweet* dan data konten *uniform resource locator*. Kedua jenis data tersebut dibandingkan dan dikelompokkan dengan algoritme *hierarchical clustering* oleh Denatari (2015), algoritme tersebut bertujuan membuat hierarki dari *cluster* dan mendapatkan *cluster* terbaik. Hasil dari algortime tersebut dapat divisualisasikan dalam bentuk dendrogram.

Untuk mendapatkan informasi sebaran geografis topik pertanian Indonesia dari hasil *clustering* data sosial media maka digunakan teknik visualisasi. Dalam penelitian ini bentuk visualisasi dibagi menjadi tiga yaitu, (1) visualisasi peta hasil *clustering*, (2) visualisasi dendrogram, dan (3) visualisasi *wordcloud*. Dengan



visualisasi tersebut pengguna dapat melihat kecenderungan data berdasarkan kata kunci yang diinginkan sehingga mempermudah proses evaluasi kerja.

## Perumusan Masalah

Berdasarkan latar belakang di atas, perumusan masalah penelitian ini adalah bagaimana menampilkan hasil *clustering* data media sosial Twitter menggunakan *framework* Shiny pada bahasa pemrograman R.

## Tujuan Penelitian

Tujuan dari penelitian ini adalah membangun sebuah aplikasi *web* yang memvisualisasikan hasil *clustering* data media sosial Twitter dalam bentuk peta, dendrogram, dan *wordcloud*.

## Manfaat Penelitian

Manfaat penelitian ini diharapkan dapat memberikan kemudahan bagi pihak terkait sektor pertanian untuk membaca informasi hasil *clustering* data *tweet* yang divisualisasikan pada peta.

## Ruang Lingkup Penelitian

Penelitian ini dibuat dengan batasan sebagai berikut:

- 1 Data dalam penelitian ini menggunakan data teks Twitter (*tweet*) yang berkaitan dengan kasus pertanian yang sudah dipraproses sebelumnya oleh Husnayain (2015), dengan jumlah data sebanyak 51 *tweet*.
- 2 Pada penelitian ini digunakan 7 kata kunci (*keyword*) sebagai data sampel *tweet* yang merepresentasikan tentang pertanian. *Keyword* tersebut terdiri atas 5 *keyword* yang berupa objek pertanian yaitu ‘padi’, ‘jagung’, ‘kelapa sawit’, ‘bawang’, dan ‘cabai’ serta 2 *keyword* yang berupa kasus pertanian yaitu ‘puso’ dan ‘hama’.
- 3 Data *tweet* tersebut kemudian dikelompokkan untuk menghitung kemiripan (*similarity*) antardata dan memberi label untuk visualisasi pada peta. *Clustering* data *tweet* dilakukan oleh Denatari (2015).

## METODE

### Data Penelitian

Data yang digunakan dalam penelitian ini adalah data teks Twitter (*tweet*) yang berkaitan dengan kasus pertanian. Data tersebut diakuisisi dari penelitian sebelumnya oleh Pratama (2015) menggunakan API. Setelah akuisisi data dari Twitter kemudian tahap selanjutnya dilakukan praproses data *tweet* untuk mendapatkan hasil *clustering* dan *document frequency* data *tweet*. Hasil *clustering* data *tweet* digunakan untuk pembuatan modul visualisasi dendrogram dan modul visualisasi peta hasil *clustering*, sementara *document frequency* data *tweet*

digunakan untuk pembuatan modul visualisasi *wordcloud*. Pada penelitian ini hasil *clustering* data *tweet* diperoleh dari peneliti sebelumnya oleh Denatari (2015) dan hasil *document frequency* data *tweet* oleh Husnayain (2015). Pada data hasil *clustering* atribut yang digunakan terdiri atas 6 atribut yang dapat dilihat pada Tabel 1.

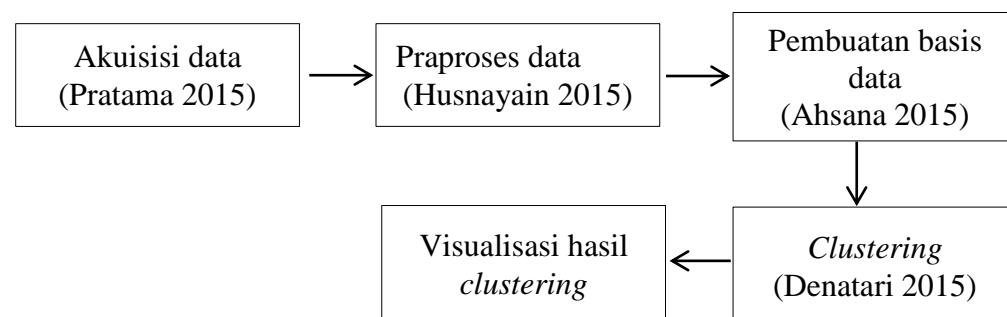
Tabel 1 Atribut data *tweet*

No	Atribut	Keterangan
1	<i>tweet</i>	Status yang dibuat pengguna
2	<i>date</i>	Waktu status dibuat
3	<i>longitude</i>	Koordinat status di garis bujur
4	<i>latitude</i>	Koordinat status di garis lintang
5	<i>cluster</i>	<i>Cluster</i> data <i>tweet</i>
6	<i>label</i>	Label data <i>tweet</i>

Enam atribut yang digunakan merupakan atribut yang dibutuhkan dalam pembuatan modul visualisasi peta hasil *clustering*. Tidak semua *tweet* yang diakuisi mengandung atribut *longitude* dan *latitude*, maka pada penelitian ini dilakukan proses pemberian nilai *longitude* dan *latitude* secara manual untuk setiap *tweet* yang tidak memiliki nilai tersebut agar lokasi *tweet* tersebut dapat terdeteksi dan dapat ditampilkan dalam peta hasil *clustering*.

### Tahapan Penelitian

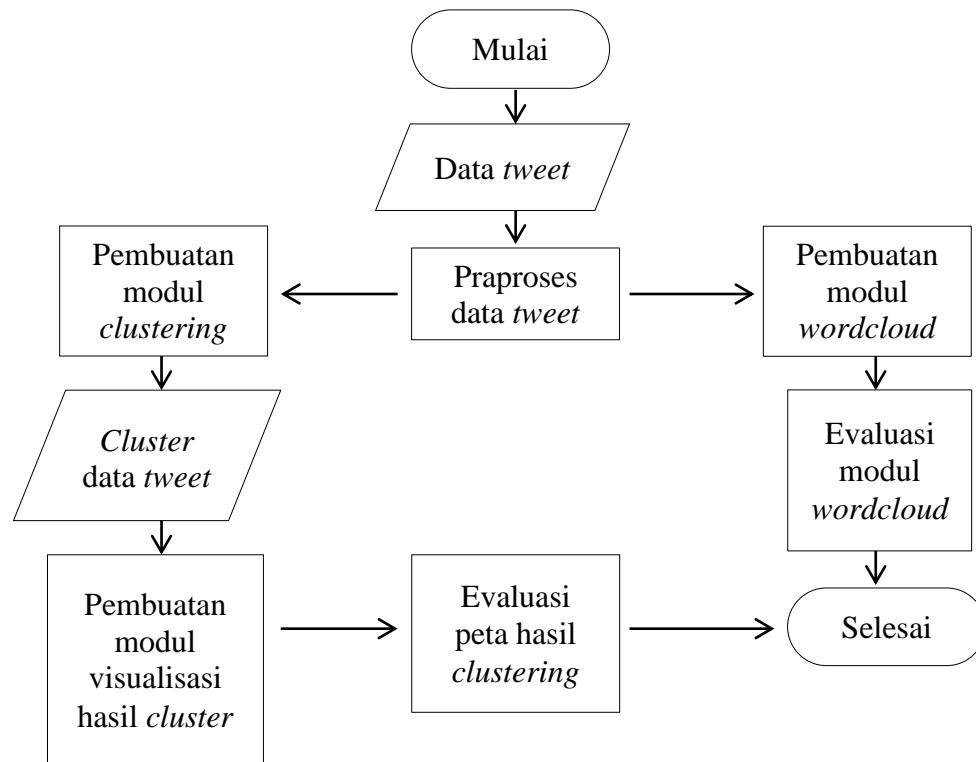
Penelitian yang dilakukan merupakan bagian dari pembangunan sistem informasi geografis berbasis media sosial yaitu difokuskan pada pembuatan modul visualisasi hasil *clustering*. Gambar 1 menunjukkan tahapan pembuatan sistem informasi geografis berbasis media sosial, sedangkan Gambar 2 menunjukkan tahapan-tahapan pembuatan modul visualisasi yang dilakukan pada penelitian ini.



Gambar 1 Gambaran umum sistem informasi geografis berbasis media sosial

Hak Cipta Dilindungi Undang-undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah  
 b. Pengutipan titik merugikan kepentingan yang wajar IPB University.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Gambar 2 Tahap penelitian

Gambar 1 menunjukkan bahwa sebelum data *tweet* divisualisasikan, data tersebut harus melewati beberapa tahapan yaitu tahap akuisisi data yang dilakukan oleh Pratama (2015), kemudian data tersebut dipraproses oleh Husnayain (2015) sebelum dimasukkan ke dalam basis data yang dibuat oleh Ahsana (2015). Data yang telah dipraproses kemudian dikelompokkan untuk mendapatkan *cluster* terbaik dan pemberian label data oleh Denatari (2015) untuk divisualisasikan.

Berdasarkan Gambar 2, tahapan penelitian ini terdiri atas proses pembuatan modul *clustering*, pembuatan modul visualisasi hasil *cluster*, dan pembuatan modul *wordcloud*. Pembuatan modul *clustering* menghasilkan *output* berupa visualisasi dalam bentuk peta dan dendrogram sedangkan pembuatan modul *wordcloud* menghasilkan *output* berupa visualisasi data teks secara visual. Data yang digunakan sebagai masukan pada aplikasi ini adalah data teks Twitter tentang kasus pertanian di Indonesia. Hasil akhir pada tahap penelitian ini adalah aplikasi hasil *clustering* data media sosial berbasis web.

### Praproses data *tweet*

Pada tahapan ini dilakukan praproses data *tweet* untuk memudahkan dalam tahapan visualisasi. Praproses data *tweet* terbagi menjadi dua bagian, bagian pertama digunakan untuk pembuatan modul *clustering* dan bagian kedua untuk pembuatan modul *wordcloud*. Data yang digunakan dalam pembuatan modul *clustering* berbeda dengan yang digunakan dalam pembuatan modul *wordcloud*. Data yang digunakan untuk pembuatan modul *clustering* mengacu pada Denatari (2015) sedangkan data yang digunakan untuk pembuatan modul *wordcloud* mengacu pada Husnayain (2015).

## Pembuatan modul *clustering*

Tahapan ini mengelompokkan data menggunakan algoritme *hierarchical clustering*. Data masukan yang digunakan adalah data *tweet* dan data konten URL. Algoritme tersebut bertujuan membuat hierarki dari *cluster* dan mendapatkan *cluster* terbaik untuk di visualisasikan ke dalam bentuk dendrogram.

## Pembuatan modul visualisasi hasil *cluster*

Pada tahapan ini menganalisis hasil *clustering* dari dendrogram yang terbentuk untuk melihat *term* dominan dan relevan pada masing-masing *cluster* sebagai dasar pemberian label pada data *tweet* untuk ditampilkan dalam peta hasil *clustering*. Terdapat beberapa parameter *linkage* dalam *hierarchical clustering*, yaitu *complete linkage*, *single linkage*, dan *average linkage*.

## Evaluasi peta hasil *clustering*

Tahap ini merupakan hasil visualiasi dalam bentuk peta berdasarkan analisis hasil *clustering* dengan melihat *term* yang dominan dan relevan. Jumlah data yang diberi label pada peta berjumlah 51 data yang terbagi ke dalam 11 *cluster*. Pada penelitian ini visualisasi pada peta menampilkan lokasi *tweet* tersebut diakses dan data hasil *clustering* berupa pola warna titik yang berbeda untuk setiap *cluster*.

## Pembuatan modul *wordcloud*

Tahap ini memasukkan data hasil praproses berupa *document frequency* dari seluruh *tweet*. *Document frequency* diambil secara acak untuk mengevaluasi jumlah kata unik pada dokumen yang dihasilkan untuk divisualisasikan dalam bentuk data teks secara visual.

## Evaluasi modul *wordcloud*

Tahapan ini merupakan hasil visualisasi berupa *wordcloud* berdasarkan hasil *document frequency* pada data *tweet*. *Wordcloud* menjelaskan kata yang dominan dari data *tweet*, semakin besar ukuran kata yang dihasilkan pada *wordcloud* maka semakin sering kata tersebut muncul pada data *tweet*.

## Lingkungan Pengembangan

Spesifikasi perangkat keras dan perangkat lunak yang digunakan untuk penelitian ini adalah sebagai berikut:

Perangkat keras berupa personal komputer dengan spesifikasi:

- Processor Intel Dual Core Pentium 2020M 2.4GHz
- RAM 2 GB
- Harddisk 500 GB SATA

Perangkat lunak:

- Bahasa pemrograman R versi 3.1.3
- RStudio-0.98.1103 dengan *package* Shiny sebagai *framework*
- Microsoft Excel 2010 untuk menyimpan data dengan format .csv



## HASIL DAN PEMBAHASAN

### Data Tweet

Data yang digunakan dalam penelitian ini adalah data teks Twitter (*tweet*) yang berkaitan dengan kasus pertanian berjumlah 51 *tweet*. Pada penelitian ini digunakan 7 kata kunci (*keyword*) sebagai data sampel *tweet* yang merepresentasikan tentang pertanian. *Keyword* tersebut terdiri atas 5 *keyword* berupa objek pertanian yaitu ‘padi’, ‘jagung’, ‘kelapa sawit’, ‘bawang’, dan ‘cabai’ serta 2 *keyword* berupa kasus pertanian yaitu ‘puso’ dan ‘hama’. Data *tweet* tersebut memiliki atribut *date*, *tweet*, *longitude*, *latitude*, *cluster*, dan label. Setiap atribut memiliki peranan yang penting dalam hasil visualisasi, misalnya atribut *date* sebagai *input* untuk menampilkan *tweet* berdasarkan waktu yang ditentukan pada peta, atribut *tweet* (isi *tweet*) untuk memberikan informasi terkait kasus pertanian, atribut *longitude* dan *latitude* untuk membaca nilai koordinat dari *tweet* yang akan ditampilkan pada peta, *cluster* untuk mengelompokkan data *tweet*, dan label untuk menunjukkan label data dari setiap *cluster*. Nilai atribut yang digunakan pada penelitian ini mengacu pada penelitian yang dilakukan oleh Pratama (2015).

Data yang digunakan dalam pembuatan aplikasi ini masih menggunakan data lokal dan tidak dapat diotomatisasi dari *database* dikarenakan modul praproses data, modul pembuatan basis data, dan modul *clustering* data yang dilakukan oleh peneliti sebelumnya masih dilakukan secara manual belum dalam bentuk aplikasi, oleh karena itu data yang digunakan tidak dapat diotomatisasi.

### Praproses Data Tweet

Praproses data *tweet* terbagi menjadi dua bagian, bagian pertama digunakan untuk pembuatan modul *clustering* dan kedua untuk pembuatan modul *wordcloud*. Dalam pembuatan modul *clustering*, tahapan praproses data yang dilakukan berupa *input term document matrix* (TDM) kemudian dilakukan perhitungan *cosine similarity* menggunakan metode *hierarchical clustering* sehingga dapat dilakukan analisis hasil *clustering*. Pada penelitian yang dilakukan oleh Pratama (2015) tidak semua *tweet* mengandung atribut *longitude* dan *latitude*, hal tersebut diakibatkan karena pada saat pengguna mengakses Twitter, GPS pada *smartphone* yang digunakan untuk mengakses *tweet* tersebut tidak dinyalakan.

*Longitude* dan *latitude* adalah suatu sistem koordinat geografis yang digunakan untuk menentukan suatu daerah di permukaan bumi. Apabila *tweet* yang akan ditampilkan pada peta tidak mengandung atribut *longitude* dan *latitude* maka lokasi *tweet* tersebut tidak dapat terdeteksi, oleh karena itu pada penelitian ini dilakukan pengambilan nilai *longitude* dan *latitude* secara manual dengan cara mengambil titik lokasi yang ada pada konten URL *tweet*. Tahapan yang dilakukan untuk mendapatkan nilai *longitude* dan *latitude* antara lain adalah dengan mencari lokasi kejadian yang ada pada *link* konten URL *tweet*, setelah mendapatkan detail lokasi pada *tweet* kemudian lokasi tersebut dimasukkan ke dalam *Google Maps* untuk mengetahui nilai *longitude* dan *latitude* lokasi pada data *tweet*. Nilai tersebut menjadi acuan untuk menunjukkan lokasi dari informasi *tweet* yang ditampilkan pada peta yang akan memudahkan pengguna dalam menyerap

informasi. Tabel 2 memperlihatkan hasil praproses data yang digunakan untuk pembuatan modul *clustering*.

Tabel 2 Hasil praproses data *clustering*

Cluster	Jumlah tweet	Label
2	3	Ancaman gagal panen
3	4	Serangan hama kelinci pada tanaman
4	1	Pemberantasan hama pada tanaman
5	14	Serangan hama keong pada tanaman padi
6	2	Tanaman hias
7	6	Penyelundupan komoditas secara illegal
8	1	Tanaman cabai
9	3	Bibit pohon kelapa
10	2	Pencegahan hama pada tanaman hias
11	2	Serangan hama pada tanaman cokelat
12	13	Gagal panen

Berdasarkan Tabel 2, jumlah data yang diberi label berjumlah 51 data yang terbagi ke dalam 11 *cluster*. Pemberian label pada data dilakukan secara subjektif dengan melihat *term* yang dominan dan relevan. Sementara untuk pembuatan modul *wordcloud*, tahapan praproses data yang dilakukan meliputi normalisasi teks, evaluasi normalisasi teks, penghapusan *stopword*, *stemming*, dan pembuatan TDM. Data hasil praproses yang digunakan dalam pembuatan modul *wordcloud* dapat dilihat pada Lampiran 2.

Dari 19 atribut yang diperoleh dari sistem pengambilan data Twitter, hanya 4 atribut yang digunakan untuk visualisasi dan 2 atribut tambahan yang didapat dari hasil praproses data *clustering* yaitu atribut *cluster* dan *label*. Contoh data *tweet* yang digunakan dalam visualisasi hasil *clustering* dari hasil pengambilan data dan hasil *clustering* data dapat dilihat pada Gambar 3.

1	tweet	date	longitude	latitude	cluster	label
2	alhamduli	9/5/2015	110.4243	-7.87304	8	Tanaman cabai
3	#Pertanian	8/5/2015	110.9621	-7.39228	12	Gagal panen
4	15 Hektare	8/5/2015	110.923	-7.38259	12	Gagal panen

Gambar 3 Contoh data *tweet* yang digunakan dalam visualisasi

Pada tahap ini, data *tweet* yang tidak mengandung atribut *longitude* dan *latitude* di dalamnya akan ditambahkan secara manual. Hal ini dilakukan untuk memberikan koordinat status di garis lintang dan bujur pada data *tweet* agar lokasi *tweet* tersebut dapat ditampilkan pada visualisasi peta hasil *clustering*.



## Pembuatan Modul *Clustering*

Pada tahap pembuatan modul *clustering*, data yang digunakan mengacu pada penelitian sebelumnya oleh Denatari (2015). Data tersebut sebagai dasar pembuatan visualisasi dalam bentuk dendrogram dan visualisasi peta hasil *clustering*. Pada visualisasi dendrogram tahapan yang dilakukan meliputi perhitungan *cosine similarity* antar dokumen yang menghasilkan nilai *cosine distance matrix*, nilai tersebut digunakan untuk pembuatan dendrogram. Tahapan selanjutnya yaitu analisis hasil *clustering*, hasil dendrogram yang terbentuk di analisis menggunakan dua pendekatan untuk melihat *term* dominan dan relevan pada masing-masing *cluster* sebagai dasar pemberian label pada data.

### *Cluster Data Tweet*

Tahap ini merupakan hasil visualisasi berupa dendrogram terhadap hasil perhitungan *cosine similarity* antar data yang dilakukan Denatari (2015). Data yang dikelompokkan menggunakan algoritme *hierarchical clustering* yang memiliki beberapa *linkage*, yaitu *complete linkage*, *single linkage*, dan *average linkage*. Masukan pada aplikasi ini berupa *term document matrix* berupa fail .RData yang berisi keterangan data *tweet* dan daftar *term* yang terdapat pada data *tweet*. Potongan kode program pada fail server dan antarmuka dapat dilihat pada Gambar 4.

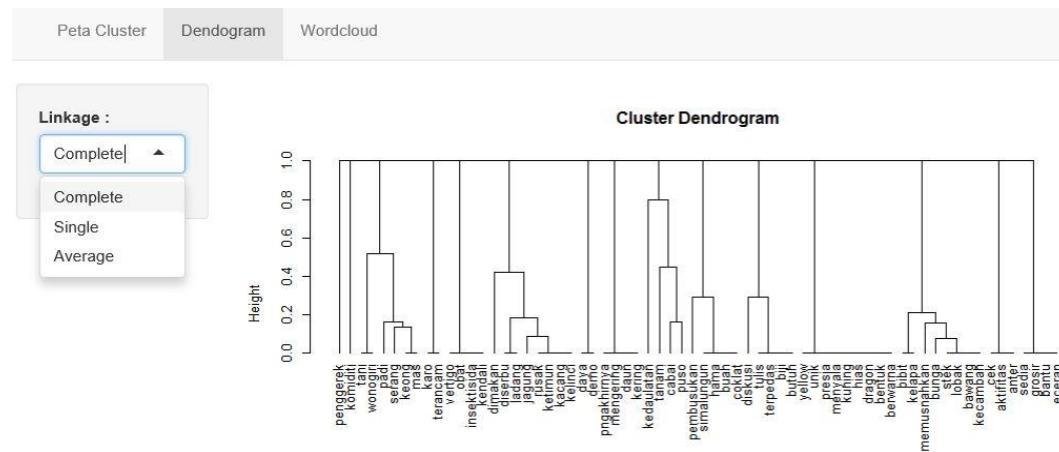
```
# load_data
load("cosine_matrix.RData")
# plot_dendrogram
switch(input$linkage,
       complete = fit <-
         hclust(dist(temp,
                     method="cosine"),
                     method="complete"),
       single   = fit <-
         hclust(dist(temp,
                     method="cosine"),
                     method="single"),
       average  = fit <-
         hclust(dist(temp,
                     method="cosine"),
                     method="average"))
plot(fit,hang=-1)
```

```
tabPanel("Dendrogram",
         sidebarPanel(width="2",
                     selectInput("linkage",
                                 "Linkage :",
                                 choices
                                 =c(
                                   "Complete"= "complete",
                                   "Single" = "single",
                                   "Average" = "average")))
         ),
```

Gambar 4 Potongan program server (kiri) dan antarmuka (kanan)

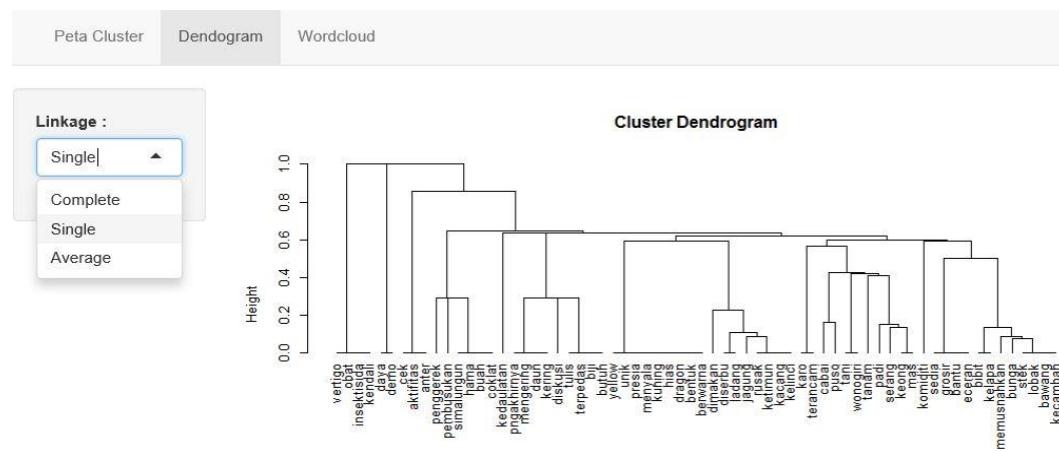
Bagian antarmuka pada aplikasi berisi instruksi-instruksi untuk menampilkan hasil *clustering* berdasarkan tipe *linkage* pada halaman *web browser* dan disimpan dalam fail UI.R. Bagian antarmuka menggunakan fungsi *fluidPage* untuk menyesuaikan ukuran *browser* pengguna. Komponen antarmuka terdiri atas *headerPanel*, *sidebarPanel*, dan *mainPanel*. *headerPanel* merupakan judul halaman aplikasi. *sidebarPanel* merupakan area yang berisi berbagai jenis masukan yang digunakan oleh pengguna.

Pada bagian server terlihat kode program dengan *method hclust*. *Method hclust* pada visualisasi ini merupakan fungsi untuk melakukan *clustering* dengan algoritme *hierarchical clustering* dan berfungsi *reactive* (reaktif). Untuk melihat hasil *clustering* yang terbentuk, pengguna harus memilih tipe *linkage* yang diinginkan pada bagian *sidebarPanel* terlebih dahulu. Pada Gambar 5 merupakan tampilan dendrogram tipe *complete linkage*.



Gambar 5 Dendrogram tipe *complete linkage*

Pada Gambar 5 dapat dilihat bahwa data telah disederhanakan dalam bentuk dendrogram dengan tipe *complete linkage* yang dikelompokkan berdasarkan kesamaan karakteristik yang ada pada dokumen. Dendrogram *complete linkage* untuk data *tweet* mengandung 18 *cluster* pada ketinggian 0.0. Gambar 6 merupakan tampilan dendrogram *single linkage*.



Gambar 6 Dendrogram tipe *single linkage*

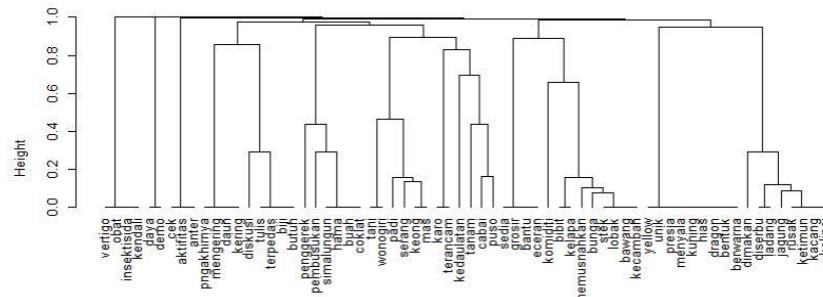
Pada Gambar 6 dapat dilihat bahwa data telah disederhanakan dalam bentuk dendrogram dengan tipe *single linkage* yang dikelompokkan berdasarkan kesamaan karakteristik yang ada pada dokumen. Dendrogram *single linkage* untuk data *tweet* mengandung 18 *cluster* pada ketinggian 0.0. Gambar 7 merupakan tampilan dendrogram *average linkage*.



Linkage :

Average  
Complete  
Single  
Average

Cluster Dendrogram

Gambar 7 Dendrogram tipe *average linkage*

Pada Gambar 7 dapat dilihat bahwa data telah disederhanakan dalam bentuk dendrogram dengan tipe *average linkage* yang dikelompokkan berdasarkan kesamaan karakteristik yang ada pada dokumen. Dendrogram *average linkage* untuk data *tweet* mengandung 18 *cluster* pada ketinggian 0.0.

Berdasarkan Gambar 5, 6, dan 7, hasil *cluster* yang terbentuk untuk ketiga parameter *linkage* memiliki kesamaan pada masing-masing anggota *cluster*. Hal tersebut dikarenakan pada saat pembuatan dendrogram, data yang dikelompokkan lebih awal adalah data yang memiliki nilai *similarity* terbesar. Terdapat perbedaan bentuk *cluster* yang terbentuk dari ketiga tipe dendrogram tersebut. Pada tipe *complete linkage* menghasilkan jarak *cluster* yang saling berjauhan satu sama lain, kemudian pada dendrogram tipe *single linkage* menghasilkan jarak *cluster* yang saling berdekatan satu sama lain sementara pada dendrogram tipe *average linkage* menggunakan prinsip nilai rata-rata dalam penentuan jarak antar *cluster*. Manfaat yang dihasilkan dari visualisasi *cluster* dalam bentuk dendrogram ini yaitu dapat diketahui *cluster* yang terbentuk, setiap *cluster* memiliki anggota masing-masing sesuai dengan kesamaan karakteristik pada dokumen.

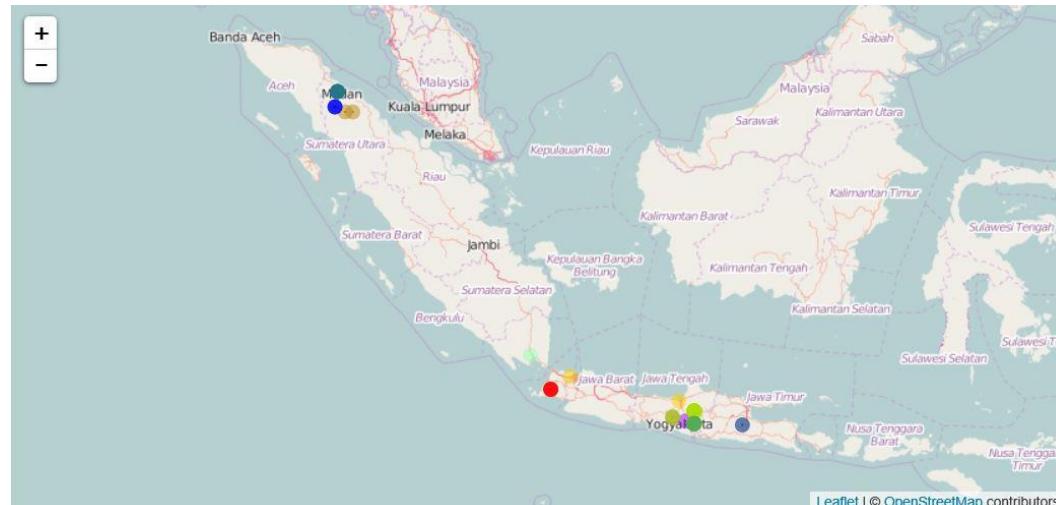
### Pembuatan Modul Visualisasi Hasil *Cluster*

Pada tahapan pembuatan modul visualisasi hasil *cluster* data yang digunakan adalah data analisis hasil *clustering* dari dendrogram yang terbentuk. Jumlah data yang digunakan pada pembuatan modul visualisasi hasil *cluster* berjumlah 51 data yang terbagi ke dalam 11 *cluster*. Pemberian label dan *cluster* pada data dilakukan secara manual berdasarkan penelitian yang dilakukan oleh Denatari (2015). Data tersebut nantinya akan digunakan sebagai dasar pembuatan peta hasil *clustering*.

### Evaluasi Peta Hasil *Clustering*

Tahap ini merupakan hasil visualisasi pada peta terhadap hasil *clustering* yang dilakukan oleh Denatari (2015) menggunakan pola warna titik. Hasil visualisasi pada peta memerlukan memerlukan 2 parameter utama, yaitu *longitude* dan *latitude*. Nilai *longitude* dan *latitude* sangat berpengaruh pada hasil visualisasi pada peta yang ditampilkan, karena apabila pada data *tweet* tidak

memiliki nilai *longitude* dan *latitude* maka lokasi *tweet* tersebut diakses tidak dapat ditampilkan di dalam visualisasi pada peta. Gambar 8 merupakan tampilan hasil visualisasi *clustering* data untuk kasus pertanian di Indonesia dengan jumlah data sebanyak 51 *tweet*.



Gambar 8 Hasil visualisasi pada peta

Pada Gambar 8 terlihat hasil visualisasi *clustering* data pada peta menggunakan pola warna titik. Titik pada peta menandakan lokasi *tweet* tersebut diakses oleh pengguna. Titik tersebut diperoleh dari nilai *longitude* dan *latitude* setiap *tweet* yang menggambarkan nilai koordinat pada GPS. Karena visualisasi ini menggunakan peta dinamis dari Leaflet yang terintegrasi langsung dengan OpenStreetMap, maka atribut *longitude* dan *latitude* sangat berperan penting untuk menampilkan *tweet* pada peta secara *online*. Tidak semua *tweet* pada data mengandung *attribute longitude* dan *latitude*, karena tidak semua pengguna mengaktifkan GPS-nya pada saat menuliskan *tweet*. Hal tersebut yang membuat penelitian ini mengambil secara langsung titik lokasi dari data *tweet* konten URL tersebut untuk ditampilkan pada visualisasi peta.

Pola warna titik pada peta memiliki hasil yang berbeda-beda, warna titik pada peta menandakan *cluster* yang terbentuk dalam *tweet*. Dari titik pada peta pengguna akan mendapatkan informasi berupa hasil *clustering* data sosial media Twitter berupa *pop up* yang berisi *tweet* tentang kasus pertanian yang dilengkapi dengan alamat URL dan *cluster* pada *tweet* tersebut. Masing-masing warna pada peta memiliki keterangan atau label yang berbeda misalnya warna biru untuk *cluster* 2 menjelaskan *tweet* tersebut tentang gagal panen, warna ungu untuk *cluster* 8 menjelaskan *tweet* tersebut tentang tanaman hias. Gambar 9 merupakan tampilan *pop up* yang berisi informasi dari *tweet* jika titik pada peta diklik.



Gambar 9 Tampilan *pop up* yang berisi informasi dari *tweet*

Untuk menampilkan sebuah visualisasi pada Gambar 9 dibutuhkan *library* Leaflet. Leaflet merupakan *package* dalam R untuk menampilkan peta secara *online* dan memungkinkan untuk menampilkan data spasial pada peta interaktif di *web browser* karena mengambil keuntungan dari *opensource library* JavaScript Leaflet.js. *Package* ini memudahkan untuk mengintegrasikan dan mengontrol peta Leaflet di R karena memiliki fitur seperti *panning* interaktif atau *zooming*, susunan peta dengan *polygons*, *lines*, *markers*, dan *pop up*.

Pengguna dapat melihat visualisasi hasil *clustering* pada peta setelah memilih tanggal yang diinginkan pada bagian `sidebarPanel` di sebelah kiri layar. Visualisasi hasil *clustering* pada peta ditampilkan pada aplikasi dengan menjalankan kode program berikut yang disimpan dalam fail Server.R:

```

1 output$map      <- renderLeaflet({
2   #select_data
3   data           <- selectedData()
4   #menggabungkan warna dengan data
5   warna         <- staticColor[as.character(data$cluster) ]
6   data          <- cbind(data,warna)
7   #atribute yang ditampilkan dalam pop up
8   popOut        <- paste("<strong>Tweet:</strong>",
9                         data$tweet,"<br><strong>Cluster:</strong>",
10                        data$cluster)
11
12   #output_maptile
13   m             <- leaflet(data) %>%addTiles('http://{s}.
14                           tile.osm.org/{z}/{x}/{y}.png',{
15                           attribution='&copy; <a href="http://osm.
16                           org/copyright">OpenStreetMap</a>contributors')
17
18   #input_by-date
19   selectInput("inputdata", "Date :", choices = c(
20     "Pilih data" = "nodata",
21     "5/5/2015" = "data1",
22     "6/5/2015" = "data2",
23     "7/5/2015" = "data3",
24     "8/5/2015" = "data4",
25     "9/5/2015" = "data5",
26     "10/5/2015" = "data6",
27     "11/5/2015" = "data7",
28     "12/5/2015" = "data8",
29     "13/5/2015" = "data9",
30     "14/5/2015" = "data10",
31     "15/5/2015" = "data11",
32     "16/5/2015" = "data12",
33     "17/5/2015" = "data13",
34     "18/5/2015" = "data14",
35     "19/5/2015" = "data15",
36     "20/5/2015" = "data16",
37     "21/5/2015" = "data17",
38     "22/5/2015" = "data18",
39     "23/5/2015" = "data19",
40     "24/5/2015" = "data20",
41     "25/5/2015" = "data21",
42     "26/5/2015" = "data22",
43     "27/5/2015" = "data23",
44     "28/5/2015" = "data24",
45     "29/5/2015" = "data25",
46     "30/5/2015" = "data26",
47     "31/5/2015" = "data27",
48     "32/5/2015" = "data28",
49     "33/5/2015" = "data29",
50     "34/5/2015" = "data30",
51     "35/5/2015" = "data31",
52     "36/5/2015" = "data32",
53     "37/5/2015" = "data33",
54     "38/5/2015" = "data34",
55     "39/5/2015" = "data35",
56     "40/5/2015" = "data36",
57     "41/5/2015" = "data37",
58     "42/5/2015" = "data38",
59     "43/5/2015" = "data39",
60     "44/5/2015" = "data40",
61     "45/5/2015" = "data41",
62     "46/5/2015" = "data42",
63     "47/5/2015" = "data43",
64     "48/5/2015" = "data44",
65     "49/5/2015" = "data45",
66     "50/5/2015" = "data46",
67     "51/5/2015" = "data47",
68     "52/5/2015" = "data48",
69     "53/5/2015" = "data49",
70     "54/5/2015" = "data50",
71     "55/5/2015" = "data51",
72     "56/5/2015" = "data52",
73     "57/5/2015" = "data53",
74     "58/5/2015" = "data54",
75     "59/5/2015" = "data55",
76     "60/5/2015" = "data56",
77     "61/5/2015" = "data57",
78     "62/5/2015" = "data58",
79     "63/5/2015" = "data59",
80     "64/5/2015" = "data60",
81     "65/5/2015" = "data61",
82     "66/5/2015" = "data62",
83     "67/5/2015" = "data63",
84     "68/5/2015" = "data64",
85     "69/5/2015" = "data65",
86     "70/5/2015" = "data66",
87     "71/5/2015" = "data67",
88     "72/5/2015" = "data68",
89     "73/5/2015" = "data69",
90     "74/5/2015" = "data70",
91     "75/5/2015" = "data71",
92     "76/5/2015" = "data72",
93     "77/5/2015" = "data73",
94     "78/5/2015" = "data74",
95     "79/5/2015" = "data75",
96     "80/5/2015" = "data76",
97     "81/5/2015" = "data77",
98     "82/5/2015" = "data78",
99     "83/5/2015" = "data79",
100    "84/5/2015" = "data80",
101    "85/5/2015" = "data81",
102    "86/5/2015" = "data82",
103    "87/5/2015" = "data83",
104    "88/5/2015" = "data84",
105    "89/5/2015" = "data85",
106    "90/5/2015" = "data86",
107    "91/5/2015" = "data87",
108    "92/5/2015" = "data88",
109    "93/5/2015" = "data89",
110    "94/5/2015" = "data90",
111    "95/5/2015" = "data91",
112    "96/5/2015" = "data92",
113    "97/5/2015" = "data93",
114    "98/5/2015" = "data94",
115    "99/5/2015" = "data95",
116    "100/5/2015" = "data96",
117    "101/5/2015" = "data97",
118    "102/5/2015" = "data98",
119    "103/5/2015" = "data99",
120    "104/5/2015" = "data100",
121    "105/5/2015" = "data101",
122    "106/5/2015" = "data102",
123    "107/5/2015" = "data103",
124    "108/5/2015" = "data104",
125    "109/5/2015" = "data105",
126    "110/5/2015" = "data106",
127    "111/5/2015" = "data107",
128    "112/5/2015" = "data108",
129    "113/5/2015" = "data109",
130    "114/5/2015" = "data110",
131    "115/5/2015" = "data111",
132    "116/5/2015" = "data112",
133    "117/5/2015" = "data113",
134    "118/5/2015" = "data114",
135    "119/5/2015" = "data115",
136    "120/5/2015" = "data116",
137    "121/5/2015" = "data117",
138    "122/5/2015" = "data118",
139    "123/5/2015" = "data119",
140    "124/5/2015" = "data120",
141    "125/5/2015" = "data121",
142    "126/5/2015" = "data122",
143    "127/5/2015" = "data123",
144    "128/5/2015" = "data124",
145    "129/5/2015" = "data125",
146    "130/5/2015" = "data126",
147    "131/5/2015" = "data127",
148    "132/5/2015" = "data128",
149    "133/5/2015" = "data129",
150    "134/5/2015" = "data130",
151    "135/5/2015" = "data131",
152    "136/5/2015" = "data132",
153    "137/5/2015" = "data133",
154    "138/5/2015" = "data134",
155    "139/5/2015" = "data135",
156    "140/5/2015" = "data136",
157    "141/5/2015" = "data137",
158    "142/5/2015" = "data138",
159    "143/5/2015" = "data139",
160    "144/5/2015" = "data140",
161    "145/5/2015" = "data141",
162    "146/5/2015" = "data142",
163    "147/5/2015" = "data143",
164    "148/5/2015" = "data144",
165    "149/5/2015" = "data145",
166    "150/5/2015" = "data146",
167    "151/5/2015" = "data147",
168    "152/5/2015" = "data148",
169    "153/5/2015" = "data149",
170    "154/5/2015" = "data150",
171    "155/5/2015" = "data151",
172    "156/5/2015" = "data152",
173    "157/5/2015" = "data153",
174    "158/5/2015" = "data154",
175    "159/5/2015" = "data155",
176    "160/5/2015" = "data156",
177    "161/5/2015" = "data157",
178    "162/5/2015" = "data158",
179    "163/5/2015" = "data159",
180    "164/5/2015" = "data160",
181    "165/5/2015" = "data161",
182    "166/5/2015" = "data162",
183    "167/5/2015" = "data163",
184    "168/5/2015" = "data164",
185    "169/5/2015" = "data165",
186    "170/5/2015" = "data166",
187    "171/5/2015" = "data167",
188    "172/5/2015" = "data168",
189    "173/5/2015" = "data169",
190    "174/5/2015" = "data170",
191    "175/5/2015" = "data171",
192    "176/5/2015" = "data172",
193    "177/5/2015" = "data173",
194    "178/5/2015" = "data174",
195    "179/5/2015" = "data175",
196    "180/5/2015" = "data176",
197    "181/5/2015" = "data177",
198    "182/5/2015" = "data178",
199    "183/5/2015" = "data179",
200    "184/5/2015" = "data180",
201    "185/5/2015" = "data181",
202    "186/5/2015" = "data182",
203    "187/5/2015" = "data183",
204    "188/5/2015" = "data184",
205    "189/5/2015" = "data185",
206    "190/5/2015" = "data186",
207    "191/5/2015" = "data187",
208    "192/5/2015" = "data188",
209    "193/5/2015" = "data189",
210    "194/5/2015" = "data190",
211    "195/5/2015" = "data191",
212    "196/5/2015" = "data192",
213    "197/5/2015" = "data193",
214    "198/5/2015" = "data194",
215    "199/5/2015" = "data195",
216    "200/5/2015" = "data196",
217    "201/5/2015" = "data197",
218    "202/5/2015" = "data198",
219    "203/5/2015" = "data199",
220    "204/5/2015" = "data200",
221    "205/5/2015" = "data201",
222    "206/5/2015" = "data202",
223    "207/5/2015" = "data203",
224    "208/5/2015" = "data204",
225    "209/5/2015" = "data205",
226    "210/5/2015" = "data206",
227    "211/5/2015" = "data207",
228    "212/5/2015" = "data208",
229    "213/5/2015" = "data209",
230    "214/5/2015" = "data210",
231    "215/5/2015" = "data211",
232    "216/5/2015" = "data212",
233    "217/5/2015" = "data213",
234    "218/5/2015" = "data214",
235    "219/5/2015" = "data215",
236    "220/5/2015" = "data216",
237    "221/5/2015" = "data217",
238    "222/5/2015" = "data218",
239    "223/5/2015" = "data219",
240    "224/5/2015" = "data220",
241    "225/5/2015" = "data221",
242    "226/5/2015" = "data222",
243    "227/5/2015" = "data223",
244    "228/5/2015" = "data224",
245    "229/5/2015" = "data225",
246    "230/5/2015" = "data226",
247    "231/5/2015" = "data227",
248    "232/5/2015" = "data228",
249    "233/5/2015" = "data229",
250    "234/5/2015" = "data230",
251    "235/5/2015" = "data231",
252    "236/5/2015" = "data232",
253    "237/5/2015" = "data233",
254    "238/5/2015" = "data234",
255    "239/5/2015" = "data235",
256    "240/5/2015" = "data236",
257    "241/5/2015" = "data237",
258    "242/5/2015" = "data238",
259    "243/5/2015" = "data239",
260    "244/5/2015" = "data240",
261    "245/5/2015" = "data241",
262    "246/5/2015" = "data242",
263    "247/5/2015" = "data243",
264    "248/5/2015" = "data244",
265    "249/5/2015" = "data245",
266    "250/5/2015" = "data246",
267    "251/5/2015" = "data247",
268    "252/5/2015" = "data248",
269    "253/5/2015" = "data249",
270    "254/5/2015" = "data250",
271    "255/5/2015" = "data251",
272    "256/5/2015" = "data252",
273    "257/5/2015" = "data253",
274    "258/5/2015" = "data254",
275    "259/5/2015" = "data255",
276    "260/5/2015" = "data256",
277    "261/5/2015" = "data257",
278    "262/5/2015" = "data258",
279    "263/5/2015" = "data259",
280    "264/5/2015" = "data260",
281    "265/5/2015" = "data261",
282    "266/5/2015" = "data262",
283    "267/5/2015" = "data263",
284    "268/5/2015" = "data264",
285    "269/5/2015" = "data265",
286    "270/5/2015" = "data266",
287    "271/5/2015" = "data267",
288    "272/5/2015" = "data268",
289    "273/5/2015" = "data269",
290    "274/5/2015" = "data270",
291    "275/5/2015" = "data271",
292    "276/5/2015" = "data272",
293    "277/5/2015" = "data273",
294    "278/5/2015" = "data274",
295    "279/5/2015" = "data275",
296    "280/5/2015" = "data276",
297    "281/5/2015" = "data277",
298    "282/5/2015" = "data278",
299    "283/5/2015" = "data279",
200    "284/5/2015" = "data280",
201    "285/5/2015" = "data281",
202    "286/5/2015" = "data282",
203    "287/5/2015" = "data283",
204    "288/5/2015" = "data284",
205    "289/5/2015" = "data285",
206    "290/5/2015" = "data286",
207    "291/5/2015" = "data287",
208    "292/5/2015" = "data288",
209    "293/5/2015" = "data289",
210    "294/5/2015" = "data290",
211    "295/5/2015" = "data291",
212    "296/5/2015" = "data292",
213    "297/5/2015" = "data293",
214    "298/5/2015" = "data294",
215    "299/5/2015" = "data295",
216    "300/5/2015" = "data296",
217    "301/5/2015" = "data297",
218    "302/5/2015" = "data298",
219    "303/5/2015" = "data299",
220    "304/5/2015" = "data300",
221    "305/5/2015" = "data301",
222    "306/5/2015" = "data302",
223    "307/5/2015" = "data303",
224    "308/5/2015" = "data304",
225    "309/5/2015" = "data305",
226    "310/5/2015" = "data306",
227    "311/5/2015" = "data307",
228    "312/5/2015" = "data308",
229    "313/5/2015" = "data309",
230    "314/5/2015" = "data310",
231    "315/5/2015" = "data311",
232    "316/5/2015" = "data312",
233    "317/5/2015" = "data313",
234    "318/5/2015" = "data314",
235    "319/5/2015" = "data315",
236    "320/5/2015" = "data316",
237    "321/5/2015" = "data317",
238    "322/5/2015" = "data318",
239    "323/5/2015" = "data319",
240    "324/5/2015" = "data320",
241    "325/5/2015" = "data321",
242    "326/5/2015" = "data322",
243    "327/5/2015" = "data323",
244    "328/5/2015" = "data324",
245    "329/5/2015" = "data325",
246    "330/5/2015" = "data326",
247    "331/5/2015" = "data327",
248    "332/5/2015" = "data328",
249    "333/5/2015" = "data329",
250    "334/5/2015" = "data330",
251    "335/5/2015" = "data331",
252    "336/5/2015" = "data332",
253    "337/5/2015" = "data333",
254    "338/5/2015" = "data334",
255    "339/5/2015" = "data335",
256    "340/5/2015" = "data336",
257    "341/5/2015" = "data337",
258    "342/5/2015" = "data338",
259    "343/5/2015" = "data339",
260    "344/5/2015" = "data340",
261    "345/5/2015" = "data341",
262    "346/5/2015" = "data342",
263    "347/5/2015" = "data343",
264    "348/5/2015" = "data344",
265    "349/5/2015" = "data345",
266    "350/5/2015" = "data346",
267    "351/5/2015" = "data347",
268    "352/5/2015" = "data348",
269    "353/5/2015" = "data349",
270    "354/5/2015" = "data350",
271    "355/5/2015" = "data351",
272    "356/5/2015" = "data352",
273    "357/5/2015" = "data353",
274    "358/5/2015" = "data354",
275    "359/5/2015" = "data355",
276    "360/5/2015" = "data356",
277    "361/5/2015" = "data357",
278    "362/5/2015" = "data358",
279    "363/5/2015" = "data359",
280    "364/5/2015" = "data360",
281    "365/5/2015" = "data361",
282    "366/5/2015" = "data362",
283    "367/5/2015" = "data363",
284    "368/5/2015" = "data364",
285    "369/5/2015" = "data365",
286    "370/5/2015" = "data366",
287    "371/5/2015" = "data367",
288    "372/5/2015" = "data368",
289    "373/5/2015" = "data369",
290    "374/5/2015" = "data370",
291    "375/5/2015" = "data371",
292    "376/5/2015" = "data372",
293    "377/5/2015" = "data373",
294    "378/5/2015" = "data374",
295    "379/5/2015" = "data375",
296    "380/5/2015" = "data376",
297    "381/5/2015" = "data377",
298    "382/5/2015" = "data378",
299    "383/5/2015" = "data379",
200    "384/5/2015" = "data380",
201    "385/5/2015" = "data381",
202    "386/5/2015" = "data382",
203    "387/5/2015" = "data383",
204    "388/5/2015" = "data384",
205    "389/5/2015" = "data385",
206    "390/5/2015" = "data386",
207    "391/5/2015" = "data387",
208    "392/5/2015" = "data388",
209    "393/5/2015" = "data389",
210    "394/5/2015" = "data390",
211    "395/5/2015" = "data391",
212    "396/5/2015" = "data392",
213    "397/5/2015" = "data393",
214    "398/5/2015" = "data394",
215    "399/5/2015" = "data395",
216    "400/5/2015" = "data396",
217    "401/5/2015" = "data397",
218    "402/5/2015" = "data398",
219    "403/5/2015" = "data399",
220    "404/5/2015" = "data400",
221    "405/5/2015" = "data401",
222    "406/5/2015" = "data402",
223    "407/5/2015" = "data403",
224    "408/5/2015" = "data404",
225    "409/5/2015" = "data405",
226    "410/5/2015" = "data406",
227    "411/5/2015" = "data407",
228    "412/5/2015" = "data408",
229    "413/5/2015" = "data409",
230    "414/5/2015" = "data410",
231    "415/5/2015" = "data411",
232    "416/5/2015" = "data412",
233    "417/5/2015" = "data413",
234    "418/5/2015" = "data414",
235    "419/5/2015" = "data415",
236    "420/5/2015" = "data416",
237    "421/5/2015" = "data417",
238    "422/5/2015" = "data418",
239    "423/5/2015" = "data419",
240    "424/5/2015" = "data420",
241    "425/5/2015" = "data421",
242    "426/5/2015" = "data422",
243    "427/5/2015" = "data423",
244    "428/5/2015" = "data424",
245    "429/5/2015" = "data425",
246    "430/5/2015" = "data426",
247    "431/5/2015" = "data427",
248    "432/5/2015" = "data428",
249    "433/5/2015" = "data429",
250    "434/5/2015" = "data430",
251    "435/5/2015" = "data431",
252    "436/5/2015" = "data432",
253    "437/5/2015" = "data433",
254    "438/5/2015" = "data434",
255    "439/5/2015" = "data435",
256    "440/5/2015" = "data436",
257    "441/5/2015" = "data437",
258    "442/5/2015" = "data438",
259    "443/5/2015" = "data439",
260    "444/5/2015" = "data440",
261    "445/5/2015" = "data441",
262    "446/5/2015" = "data442",
263    "447/5/2015" = "data443",
264    "448/5/2015" = "data444",
265    "449/5/2015" = "data445",
266    "450/5/2015" = "data446",
267    "451/5/2015" = "data447",
268    "452/5/2015" = "data448",
269    "453/5/2015" = "data449",
270    "454/5/2015" = "data450",
271    "455/5/2015" = "data451",
272    "456/5/20
```

```

        "10/5/2015" = "data6",
        "11/5/2015" = "data7"))

#output_popup
8 output$info <- renderPrint({
9 mydfTweet    <- selectedData()
10 tweetText   <- nearPoints(mydfTweet,
                           input$plot_hover,xvar="longitude",
                           yvar="latitude") })

```

Kode program di atas adalah kode untuk menampilkan hasil *clustering* data pada peta berdasarkan tanggal. Kode pada baris 2 adalah kode untuk membaca *input*-an data berupa *csv file* data *tweet* pertanian di Indonesia. Fungsi *staticColor* pada baris 4 dan 5 digunakan untuk pemberian warna pada titik, warna tersebut disesuaikan dengan hasil *cluster* pada data *tweet*. Kode pada baris 6 digunakan untuk menampilkan peta dengan map dasar OpenStreetMap. Kode pada baris 7 digunakan untuk melihat hasil visualisasi pada peta berdasarkan tanggal yang dipilih. Fungsi *staticColor* pada baris 10 digunakan untuk menampilkan informasi dari *tweet* berupa *pop up* yang akan muncul ketika titik pada peta diklik.

Jumlah data *tweet* yang digunakan untuk visualisasi hasil cluster data berjumlah 50 data yang terbagi ke dalam 11 *cluster*. Karena jumlah data yang terbatas maka pada menu peta cluster ini juga terdapat fungsi *sort by date* yang berfungsi untuk menampilkan informasi berdasarkan tanggal yang diinginkan oleh pengguna.

### Pembuatan Modul *Wordcloud*

Pada tahapan pembuatan modul *wordcloud* data yang digunakan mengacu pada penelitian sebelumnya oleh Husnayain (2015). Pada visualisasi *wordcloud* tahapan yang dilakukan meliputi normalisasi teks, penghapusan *stopword*, *stemming*, dan pembuatan *term document matrix* (TDM). Dari tahapan tersebut didapatkan hasil *document frequency* (DF) yang digunakan sebagai dasar pembuatan visualisasi *wordcloud*.

### Evaluasi Modul *Wordcloud*

Tahapan ini merupakan visualisasi hasil praproses data teks Twitter menggunakan *wordcloud*. Gambar 10 merupakan hasil visualisasi data teks secara visual, dengan menampilkan kata yang dominan muncul pada data *tweet*. Data frekuensi kemunculan dapat dilihat pada Lampiran 3. Menurut Zhao (2013), *wordcloud* digunakan untuk menyajikan kata-kata penting dalam dokumen. Pada akhirnya, kata-kata dan *tweet* akan mengelompok membentuk sebuah *cluster*. Keunggulan yang dihasilkan dari visualisasi data teks berupa *wordcloud* ini yaitu hasil visualisasi yang ditampilkan lebih menarik dengan melihat tata letak dan skema warna yang dihasilkan serta dapat mengetahui kata yang dominan muncul pada data *tweet* dengan melihat ukuran *font* yang dihasilkan.



Gambar 10 Hasil visualisasi modul *wordcloud*

*Wordcloud* dibuat berdasarkan hasil praproses data *tweet* yang dilakukan oleh Husnayain (2015). Praproses data *tweet* tersebut dilakukan dengan mengkombinasikan teknik-teknik praproses, yaitu normalisasi teks, *stopword removal*, dan *stemming* dengan *document frequency* (DF) *thresholding*. DF adalah jumlah dokumen yang muncul pada setiap kata atau *term*, sementara DF *thresholding* adalah teknik untuk mengurangi kosakata. Contoh hasil DF *thresholding* data *tweet* diberikan pada Tabel 3.

Tabel 3 Contoh DF pada data *tweet*

No	<i>Term</i>	DF
1	tanaman	49
2	<i>hectare</i>	27
3	cabai	15
4	sawit	6
5	komoditi	1

Berdasarkan Gambar 10, data yang ditampilkan dalam *wordcloud* berupa DF *thresholding* data *tweet*. *Term* dari hasil DF *thresholding* berjumlah 87 kata, setiap *term* memiliki nilai frekuensi yang dihasilkan. *Term* yang memiliki frekuensi paling besar hanya ada pada *term* tanaman saja dengan jumlah frekuensi 49, sementara *term* yang memiliki frekuensi paling rendah terdapat 35 kata dengan jumlah frekuensi 1. Berikut ini merupakan kode implementasi visualisasi praproses data yang disimpan dalam fail Server.R:

```

#read csv file
1 wcTweet <- as.data.frame(read.csv
  (file="tabelfreq1.csv", stringsAsFactors=FALSE));
2 #wordcloud funtion
  output$cloud <- renderPlot({
3 wordcloud(wcTweet$Word,
            wcTweet$Frekuensi,
4 min.freq = 20,
5 scale = c(4,1),
6 colors = brewer.pal(8,"Dark2"),
            random.order = FALSE,
            rot.per = 0.5)
})

```

Kode program di atas adalah kode untuk menampilkan visualisasi *wordcloud* untuk data *tweet*. Kode pada baris 1 adalah kode untuk membaca data *file csv* hasil DF *thresholding*. Fungsi *wordcloud* pada baris 2 adalah untuk membuat plot *wordcloud*. Fungsi *min.freq* pada baris 3 yaitu jika ada kata-kata dengan frekuensi di bawah minimum frekuensi maka tidak akan diplotkan. Variabel *scale* pada baris 4 digunakan untuk menunjukkan ukuran kata. Variabel *colors* pada baris 5 digunakan untuk memberikan warna pada kata yang paling dominan muncul.

Semua informasi yang di masukkan berdasarkan kata kunci pada data *tweet* yang berjumlah 51 *tweet* bisa ditampilkan di dalam *wordcloud*. Untuk mengetahui kata yang paling dominan muncul dari informasi yang ditentukan berdasarkan kata kunci sebelumnya. Kita dapat melihat pada visualisasi *wordcloud*, semakin besar ukuran *term* yang divisualisasikan dalam *wordcloud* maka semakin dominan *term* tersebut muncul pada setiap *tweet* dan semakin kecil ukuran *term* yang dihasilkan maka semakin jarang *term* tersebut muncul pada setiap *tweet*.

## Rencana Pengembangan Sistem

Aplikasi ini dibangun untuk menyempurnakan pembuatan aplikasi web sistem informasi geografis (SIG) berbasis Twitter untuk kasus pertanian di Indonesia. Aplikasi web SIG ini terdiri dari 5 modul yaitu, akuisisi data yang dilakukan oleh Pratama (2015), praproses data oleh Husnayain (2015), pembuatan basis data oleh Ahsana (2015), *clustering* oleh Denatari (2015), dan visualisasi hasil *clustering* yang dilakukan pada penelitian ini. Pada pengembangan SIG ini modul-modul yang dibuat saling terkait. Pertama yaitu modul akuisisi data untuk mempermudah pengguna dalam pengambilan data Twitter dan mempermudah modul berikutnya pada saat praproses data. Kedua yaitu modul praproses data teks Twitter dengan menerapkan metode *text mining* untuk menghilangkan kata yang tidak bermakna, tidak teratur, dan tidak baku. Ketiga yaitu modul pembuatan basis data yang bertujuan membangun basis data untuk mempermudah proses transportasi data. Keempat yaitu modul *clustering* untuk memberikan label pada data teks Twitter dengan topik pertanian sehingga bermanfaat untuk proses visualisasi data. Kelima yaitu modul visualisasi hasil *clustering* untuk membangun sebuah aplikasi web yang memvisualisasikan hasil *clustering* data sosial media Twitter dalam bentuk dendrogram, peta, dan *wordcloud*.



Pada modul akuisisi data oleh Pratama (2015) aplikasi yang dibangun dapat dilakukan secara dinamis, data yang digunakan pada penelitian tersebut bersifat *real-time*, jadi pengguna dapat mengetahui informasi yang diinginkan mengenai kasus pertanian secara *up-to-date*. Sementara pada modul praproses data oleh Husnayain (2015), modul pembuatan basis data oleh Ahsana (2015), dan modul *clustering* oleh Denatari (2015) masih dilakukan secara manual. Hal tersebut mengakibatkan pada modul visualisasi hasil *clustering*, data yang divisualisasikan pada peta tidak dapat menampilkan informasi mengenai kasus pertanian secara *up-to-date* sesuai dengan apa yang diinginkan oleh pengguna, karena data yang digunakan dalam visualisasi hasil *clustering* adalah data yang sudah melalui tahapan praproses sebelumnya oleh Husnayain (2015) dan *clustering* oleh Denatari (2015). Jadi aplikasi web sistem informasi geografis (SIG) berbasis Twitter untuk kasus pertanian di Indonesia ini dapat diakses oleh pengguna apabila setiap modul yang dikerjakan dapat dilakukan secara dinamis atau otomatis agar pada proses visualisasi hanya memanggil *query* yang diinginkan pada *database* yang dibuat oleh Ahsana (2015).

### Kelemahan Aplikasi

Pada aplikasi visualisasi hasil *clustering* ini terdapat 51 *tweet* yang berisi informasi tentang kasus pertanian dan dapat ditampilkan pada peta berdasarkan tanggal. Salah satu atribut yang digunakan untuk menampilkan data hasil *clustering* pada peta yaitu *longitude* dan *latitude*, tanpa atribut tersebut maka *tweet* yang berisikan informasi tentang kasus pertanian tidak dapat ditampilkan pada peta. Aplikasi ini juga dapat menampilkan informasi berupa data teks secara visual dalam bentuk *wordcloud*, berdasarkan keluaran dari *wordcloud* pengguna dapat mengetahui kata yang dominan muncul pada data *tweet* dengan melihat ukuran *font* yang dihasilkan.

Kelemahan pada aplikasi visualisasi hasil *clustering* ini yaitu informasi yang ditampilkan pada peta masih terbatas. Pada aplikasi ini masih menggunakan data yang dipraproses secara manual dan statis. Untuk mendapatkan informasi yang *up-to-date* sesuai dengan apa yang diinginkan oleh pengguna maka pada penelitian ini harus menggunakan data yang melalui tahapan praproses secara dinamis yang telah dimasukkan dalam sebuah basis data.

## SIMPULAN DAN SARAN

### Simpulan

Kesimpulan dari hasil penelitian ini, yaitu visualisasi hasil *clustering* data berbasis media sosial untuk kasus pertanian di Indonesia berbasis web telah berhasil dibangun dengan menggunakan *framework* Shiny. Data media sosial yang digunakan adalah data teks Twitter yang berjumlah 51 *tweet*. Fitur utama yang terdapat pada aplikasi ini adalah peta *cluster*, pengguna dapat melihat hasilnya yang berupa sebaran informasi yang berkaitan dengan kasus pertanian yang divisualisasikan dalam bentuk peta. Fitur lainnya adalah pengguna dapat melihat hasil visualisasi dalam bentuk dendrogram menggunakan algoritme

*hierarchical clustering*, pengguna dapat mengetahui *cluster* yang terbentuk berdasarkan kemiripan karakteristik pada dokumen. Selain itu, aplikasi ini juga menampilkan hasil visualisasi data teks Twitter dalam bentuk *wordcloud*, pengguna dapat mengetahui kata yang dominan muncul pada setiap *tweet*.

### Saran

Saran untuk penelitian dan pengembangan aplikasi selanjutnya sebagai berikut:

- 1 Aplikasi dikembangkan dengan menggunakan teknik visualisasi yang lain selain menggunakan *library* Leaflet untuk menghasilkan visualisasi pada peta yang lebih interaktif dan informatif.
- 2 Teknik visualisasi berikutnya diharapkan juga menampilkan data statistik pertanian, sehingga pengguna dapat membandingkan data *tweet* dengan data statistik pertanian.
- 3 Visualisasi yang dihasilkan dapat dilakukan secara dinamis, karena pada penelitian ini tidak semua visualisasi yang dihasilkan dapat dilakukan secara dinamis.
- 4 Visualisasi yang ditampilkan harus lebih spesifik dan beragam tergantung teknik pengolahan datanya.

### DAFTAR PUSTAKA

- Ahsana LN. 2015. Modul integrasi R-Studio dan MongoDB untuk sistem informasi geografis pertanian indonesia [skripsi]. Bogor (ID): Institut Pertanian Bogor.
- Denatari RB. 2015. Clustering data teks Twitter untuk kasus pertanian di Indonesia [skripsi]. Bogor (ID): Institut Pertanian Bogor.
- Husnayain I. 2015. Praproses data teks Twitter pada sistem informasi geografis untuk kasus pertanian di Indonesia [skripsi]. Bogor (ID): Institut Pertanian Bogor.
- [Kemenkominfo]. 2013. Kominfo: internet hanya digunakan untuk sekedar *update* status [Internet]. [diunduh 2014 Nov 24]. Tersedia pada: <http://kominfo.go.id/index.php>.
- Pratama FS. 2015. Akuisisi data pada sistem informasi geografis berbasis Twitter untuk kasus pertanian di Indonesia [skripsi]. Bogor (ID): Institut Pertanian Bogor.
- Sugianto. 2010. Sistem informasi geografis untuk pemetaan dan analisa daerah pertanian di Kabupaten Ponorogo [skripsi]. Surabaya (ID): Institut Teknologi Sepuluh Nopember.
- Susanto H. 2014. Visualisasi data teks Twitter berbasis Bahasa Indonesia menggunakan teknik pengklasteran [skripsi]. Surabaya (ID): Institut Teknologi Sepuluh Nopember.
- Zhao Y. 2013. R and Data Mining: Example and Case Studies Academic. Elsevier. hlm 105-106.



## LAMPIRAN

Lampiran 1 Contoh data *tweet* tentang pertanian di Indonesia

<i>tweet</i>		<i>date</i>	<i>longitude</i>	<i>latitude</i>	<i>cluster</i>	<i>label</i>
alhamdulillah sudah sampai rmh lagi, abis anter sekolah anak, selanjutnya aktifitas pagi cek tanaman cabai... http://t.co/XGEOXqLjOX		9/5/2015	110.4242874	-7.873044	8	Tanaman cabai
#Pertanian Puluhan Tanaman Cabai Kedaulatan Rakyat http://t.co/pFNZjGVKoq	Gawat! Hektare Puso -	8/5/2015	110.9621404	-7.392276	12	Gagal panen
15 Hektare Cabai Tanaman Cabai Puso http://t.co/H5lTEJeYxO	Tanaman Puso -	8/5/2015	110.922965	-7.382589	12	Gagal panen
15 Hektare Cabai Tanaman Cabai Puso http://t.co/S2GZNVexd0	Tanaman Puso -	8/5/2015	110.922965	-7.382589	12	Gagal panen
15 Hektare Cabai Tanaman Cabai Puso http://t.co/D89EEcYZ67	Tanaman Puso -	8/5/2015	110.922965	-7.382589	12	Gagal panen
15 Hektare Cabai Tanaman Cabai Puso http://t.co/ZdQBaB3AfS	Tanaman Puso -	8/5/2015	110.922965	-7.382589	12	Gagal panen
15 Hektare Cabai Tanaman Cabai Puso http://t.co/ova3apUnJe	Tanaman Puso -	8/5/2015	110.922965	-7.382589	12	Gagal panen
Insektisida Vertigo 100 Ec - Rp 85,000: Insektisida vertigo 100EC merupakan obat yang mampu mengendalikan ham... http://t.co/HEr1IYc8Po	100 Ec - Rp 85,000: Insektisida vertigo 100EC merupakan obat yang mampu mengendalikan ham...	10/5/2015	111.8960479	-7.156088	6	Tanaman hias
Jangan Hanya Bisa Demo saja Minta ini itu tanpa ada daya upaya,, #AyooMenanam ^_^... http://t.co/2U7Haro1jR	Jangan Hanya Bisa Demo saja Minta ini itu tanpa ada daya upaya,, #AyooMenanam ^_^...	10/5/2015	105.200697	-5.428404	4	Pemberantasan hama pada tanaman

Hak Cipta Dilindungi Undang-undang  
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah  
b. Pengutipan titik merugikan kepentingan yang wajar IPB University.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

### Lampiran 1 Lanjutan

<i>tweet</i>	<i>date</i>	<i>longitude</i>	<i>latitude</i>	<i>cluster</i>	<i>label</i>
RT @InfoSimalungun: Tanaman Coklat Simalungun Diserang Hama Pembusukan Buah http://t.co/YmHbSjKIW Z #Simalungun http://t.co/orag0t87Ro	10/5/2015	99.067711	2.9559254	11	Serangan hama pada tanaman coklat
Tanaman Coklat di Simalungun Diserang Hama Penggerek Buah http://t.co/3NbNjCEI1F #eksposnews	9/5/2015	98.769051	2.9519263	11	Serangan hama pada tanaman coklat
RT @karo_news: Ratusan Hektar Tanaman Padi di Karo Terancam Puso http://t.co/g5DP38Phgi	6/5/2015	98.4559585	3.1119222	2	Ancaman gagal panen
Ratusan Ha Tanaman Padi di Karo Terancam Puso http://t.co/C9x0ltds3j	5/5/2015	98.4559585	3.1120074	2	Ancaman gagal panen
PERTANIAN WONOGIRI : Keong Mas Serang 25 Hektare Tanaman Padi => http://t.co/mCfZDy8wp W	11/5/2015	110.862044	-7.783713	5	Serangan hama keong pada tanaman padi
RT @pertanian_ID: #Pertanian Keong Mas Serang 25 Hektare Tanaman Padi - Solopos http://t.co/uhcOoESYd6	10/5/2015	110.862044	-7.783713	5	Serangan hama keong pada tanaman padi
Selain bawang, petugas juga memusnahkan 30 kg kecambah kelapa sawit, bibit lobak dan stek tanaman bunga yang d... http://t.co/jgh0r5JL3E	7/5/2015	98.5294049	3.6426183	7	Penyelundupan komoditas secara ilegal



Hak Cipta Dilindungi Undang-undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.  
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

## Lampiran 2 Data hasil frekuensi dokumen

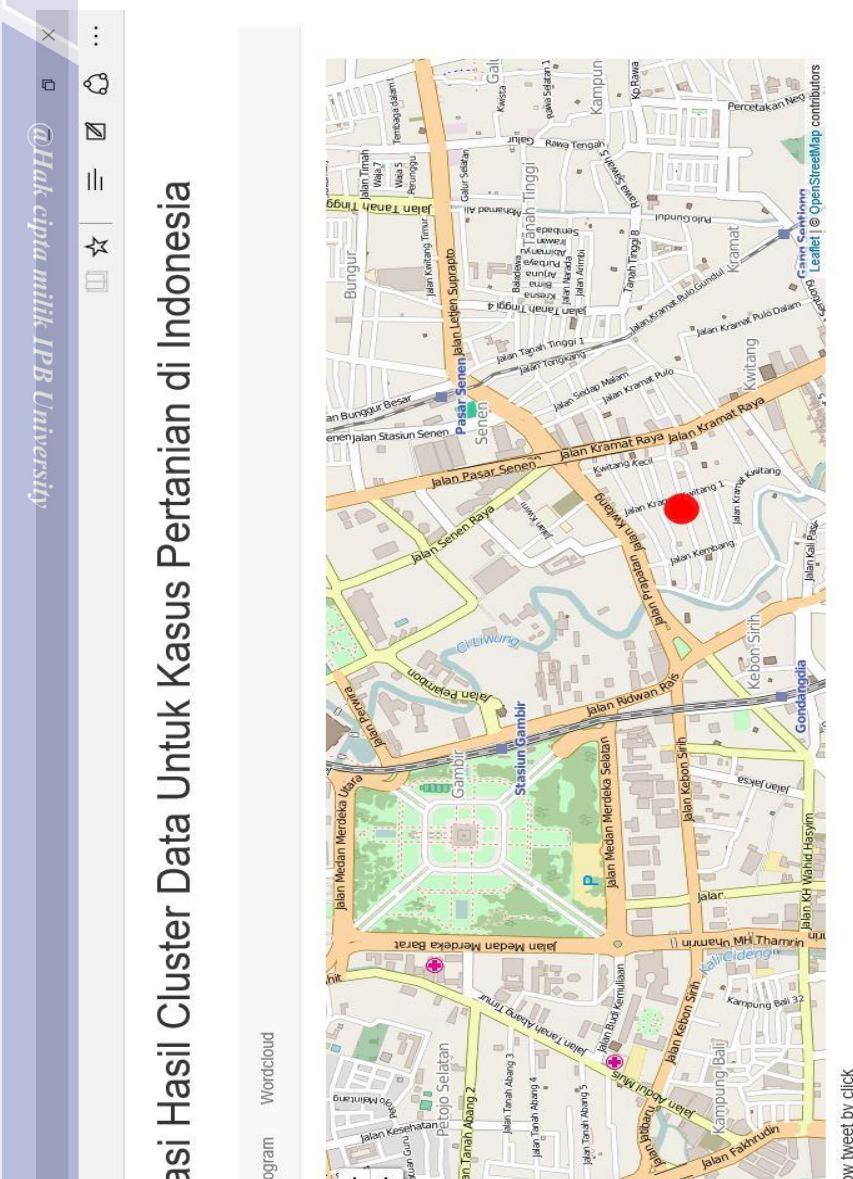
No	Term	Frekuensi dokumen
1	tanaman	49
2	hectare	27
3	padi	17
4	puso	16
5	cabai	15
6	serang	14
7	keong	12
8	bibit	8
9	kelapa	8
10	bunga	7
11	bawang	6
12	jagung	6
13	kecambah	6
14	lobak	6
15	puluhan	6
16	sawit	6
17	stek	6
18	hutan	5
19	kacang	5
20	kelinci	5
21	ketimun	5
22	memusnahkan	5
23	petugas	5
24	rusak	5
25	tanah	5
26	diserbu	4
27	ladang	4
28	pertanian	4
29	warga	4
30	wonogiri	4
31	dimakan	3
32	karo	3
33	ratusan	3
34	terancam	3
35	bantu	2
36	buah	2
37		

### Lampiran 3 Kode program untuk menampilkan plot berdasarkan tanggal

```
#read csv file
1 "<-(dfdata1,
  as.data.frame(read.csv(file="tweet_5mei.csv",
  stringsAsFactors=FALSE)))
2 "<-(dfdata2,
  as.data.frame(read.csv(file="tweet_6mei.csv",
  stringsAsFactors=FALSE)))
3 "<-(dfdata3,
  as.data.frame(read.csv(file="tweet_7mei.csv",
  stringsAsFactors=FALSE)))
4 "<-(dfdata4,
  as.data.frame(read.csv(file="tweet_8mei.csv",
  stringsAsFactors=FALSE)))
5 "<-(dfdata5,
  as.data.frame(read.csv(file="tweet_9mei.csv",
  stringsAsFactors=FALSE)))
6 "<-(dfdata6,
  as.data.frame(read.csv(file="tweet_10mei.csv",
  stringsAsFactors=FALSE)))
7 "<-(dfdata7,
  as.data.frame(read.csv(file="tweet_11mei.csv",
  stringsAsFactors=FALSE)))
8 "<-(dfalldata,
  as.data.frame(read.csv(file="data10.csv",
  stringsAsFactors=FALSE)))

selectedData <- reactive ({
  switch(input$inputdata,
    "data1" = {"<-(df_tweet, dfdata1)" },
    "data2" = {"<-(df_tweet, dfdata2)" },
    "data3" = {"<-(df_tweet, dfdata3)" },
    "data4" = {"<-(df_tweet, dfdata4)" },
    "data5" = {"<-(df_tweet, dfdata5)" },
    "data6" = {"<-(df_tweet, dfdata6)" },
    "data7" = {"<-(df_tweet, dfdata7)" },
    "nodata" = {"<-(df_tweet, dfalldata)" },
  )
})
```

## Lampiran 4 Plot hasil *clustering* data pertanian secara detail lokasi



## RIWAYAT HIDUP

Penulis dilahirkan di Jakarta pada tanggal 30 September 1993 dari pasangan Harsana dan Sri Amani. Penulis merupakan anak pertama dari tiga bersaudara. Pada tahun 2011 penulis lulus dari SMA PGRI 1 Bekasi dan pada tahun yang sama penulis lulus seleksi masuk Institut Pertanian Bogor (IPB) melalui Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) Undangan dan diterima di Departemen Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam.

Pada bulan Juli hingga Agustus tahun 2014, penulis melaksanakan kegiatan praktik kerja lapang di Bank Rakyat Indonesia dengan bidang kajian pemrograman web dan analisis *database*.

