



**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

**TUGAS AKHIR - KS 091336**

**PENERAPAN KLASIFIKASI TEKS UNTUK  
PEMETAAN DATA TWITTER PELANGGAN PT.  
TELEKOMUNIKASI SELULAR DALAM BENTUK  
*HEAT MAP***

Mochamad Nizar Palefi Ma'ady  
NRP 5210 100 074

Dosen Pembimbing I  
Prof. Ir. Arif Djunaidi M.Sc., Ph.D.

Dosen Pembimbing II  
Renny Pradina Kusumawardani S.T., M.T.

JURUSAN SISTEM INFORMASI  
Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember  
Surabaya 2014



**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

**FINAL PROJECT - KS 091336**

**APPLICATION OF TEXT CLASSIFICATION  
FOR MAPPING TWITTER DATA OF PT.  
TELEKOMUNIKASI SELULAR CUSTOMERS  
USING HEAT MAP**

Mochamad Nizar Palefi Ma'ady  
NRP 5210 100 074

Supervisors  
Prof. Ir. Arif Djunaidy M.Sc., Ph.D.  
Renny Pradina Kusumawardani S.T., M.T.

DEPARTMENT OF INFORMATION SYSTEM  
Faculty of Information Technology  
Institute of Technology Sepuluh Nopember  
Surabaya 2014

# **PENERAPAN KLASIFIKASI TEKS UNTUK PEMETAAN DATA TWITTER PELANGGAN PT. TELEKOMUNIKASI SELULAR DALAM BENTUK *HEAT MAP***

**Nama Mahasiswa : Mochamad Nizar Palefi Ma'ady**  
**NRP : 5210 100 074**  
**Jurusan : Sistem Informasi FTIF-ITS**  
**Dosen Pembimbing : Prof. Ir. Arif Djunaidy M.Sc., Ph.D.**  
**Renny Pradina Kusumawardani S.T, M.T**

## **Abstrak**

*Persaingan yang semakin sengit membuat berbagai perusahaan semakin gencar untuk memperoleh informasi dari berbagai media untuk keperluan analisis layanan yang diberikan. Data yang terekam dalam jejaring sosial twitter dapat dijadikan sebagai data yang sangat potensial untuk dianalisis karena twitter merupakan salah satu media sosial yang paling banyak penggunanya di dunia.*

*Dalam tugas akhir ini dibangun sebuah aplikasi pengklasifikasi teks berbasis Naive Bayes terhadap data twitter pelanggan PT. Telekomunikasi Selular untuk memantau opini positif dan negatif para pelanggannya terhadap layanan yang diberikan dalam bentuk heat map. Sebelum pengklasifikasi teks dibangun, sebuah praproses data dilakukan untuk menentukan atribut, memilih data yang memiliki koordinat, dan memilah kata yang dikategorikan sebagai opini positif dan negatif. Pengklasifikasi teks berbasis Naive Bayes kemudian dibangun dengan menggunakan fitur kata opini yang telah diperoleh. Koordinat data yang telah diklasifikasikan kemudian digunakan untuk memvisualisasikan gradasi tingkat kepadatan dari opini positif dan negatif yang yang diperoleh dengan menggunakan heat map.*

*Aplikasi pengklasifikasi teks diujicobakan terhadap sejumlah sampel data yang diperoleh dari twitter pada bulan*

*Desember 2013 serta data pada April dan Juni 2014 memberikan hasil akurasi dan recall yang memadai dengan nilai berturut-turut 73% dan 91%. Penggunaan heat map untuk memetakan posisi koordinat pelanggan yang opininya telah diklasifikasi akan memudahkan pihak pengguna untuk mengamati tingkat kepadatan opini pelanggan berdasarkan visualisasi gradasi warna peta yang disajikan pada radius tertentu dari satu titik koordinat pengamatan yang diinginkan.*

**Kata Kunci: PT. Telekomunikasi Selular, Twitter API Stream, Klasifikasi Teks, Naïve Bayes, Visualisasi, Heat Map**

**APPLICATION OF TEXT CLASSIFICATION FOR  
MAPPING TWITTER DATA OF PT. TELEKOMUNIKASI  
SELULAR CUSTOMERS USING HEAT MAP**

**Student Name** : Mochamad Nizar Palefi Ma'ady  
**SID** : 5210 100 074  
**Department** : Sistem Informasi FTIF-ITS  
**Academic Promotor** : Prof. Ir. Arif Djunaidy M.Sc., Ph.D.  
Renny Pradina Kusumawardani S.T, M.T

**Abstract**

*The tight competition makes companies increasing effort to gain information from various media for the need of service analysis that will be given. Recorded data in social media such as twitter can be used as a potential data to be analyzed because it is one of the most used social media in the world.*

*In this thesis, there will be a construction of naive bayes based text classifier application on telkomsel corporation customer data twitter to observe positive and negative opinion of the customer on the services given in the form of heat map. Before the text classifier application is built, a pre-data process will be done in order to determine attribute, to choose data that have coordinate, and to decide which words are considered as positive or negative opinion. After it is finished, a naive bayes based text classifier will be built by using word opinion feature which has been gained. Then, classified data coordinate will be used to visualize density level gradation of positive and negative opinion that have been gained from the use of heat map.*

*Text classifier application is tested and used in a number of data sample gained from twitter in the month of December 2013, April and June 2014. The test gives accurate result and proportional recall in succesion of 73% and 91%. The use of heat map to determine customer coordinate position whose opinion has been classified, will ease the user to observe customer*

*opinion density level based on colour gradation visualization which are presented in certain radius from one coordinate observing point that one want.*

***Keywords: PT. Telekomunikasi Selular, Twitter API Stream, Text Classification, Naïve Bayes, Visualization, Heat Map***

**PENERAPAN KLASIFIKASI TEKS UNTUK PEMETAAN  
DATA TWITTER PELANGGAN PT. TELEKOMUNIKASI  
SELULAR DALAM BENTUK *HEAT MAP***

**TUGAS AKHIR**

Disusun Untuk Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer  
pada

Jurusan Sistem Informasi  
Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember

Oleh:

**MOCHAMAD NIZAR PALEFI MA'ADY**  
NRP. 5210 100 074

Surabaya, Juli 2014

**Ketua Jurusan Sistem Informasi**

**Dr. Eng. FEBRILIYAN SAMOPA, S.Kom., M.Kom.**  
**NIP.19730219 199802 1 001**



**PENERAPAN KLASIFIKASI TEKS UNTUK PEMETAAN  
DATA TWITTER PELANGGAN PT. TELEKOMUNIKASI  
SELULAR DALAM BENTUK *HEAT MAP***

**TUGAS AKHIR**


Disusun Untuk Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer  
pada  
Jurusan Sistem Informasi  
Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember

Oleh :

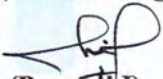
**MOCHAMAD NIZAR PALEFI MA'ADY**  
**NRP. 5210 100 074**

Disetujui Tim Penguji : Tanggal Ujian : 18 Juli 2014

Periode Wisuda : September 2014

Prof. Ir. Arif Djunaidy M.Sc., Ph.D.  (Pembimbing I)

Renny Pradina Kusumawardani S.T., M.T.  (Pembimbing II)

Mahendrawathi Erawan S.T., M.Sc., Ph.D.  (Penguji I)

Retno Aulia Vinarti S.Kom, M.Kom  (Penguji II)



## KATA PENGANTAR

Alhamdulillah, puji syukur atas kehadiran Allah SWT atas berkah, rahmat dan limpahanNya sehingga tidak hanya atas terselesaikannya buku tugas akhir ini, namun juga atas segala cerita yang diaturNya.

Tugas akhir yang berjudul “**Penerapan Klasifikasi Teks Untuk Pemetaan Data Twitter Pelanggan PT. Telekomunikasi Selular dalam Bentuk *Heat Map***” ini menjadi tonggak terakhir dari rentetan tugas-tugas, *final project*, ujian selama delapan semester masa belajar penulis di di jurusan Sistem Informasi, Fakultas Teknologi Informasi, ITS.

Semoga tugas akhir ini tidak menjadi titik pemberhentian dari perjalanan untuk terus mencari ilmu. Semoga tugas akhir ini menjadi pengantar ke dalam ruang pengetahuan yang lebih luas. Mereka yang menginspirasi, mereka yang menjadi teladan, mereka yang menjadi panutan, Bapak dan Ibu dosen Jurusan Sistem Informasi, termasuk mereka yang secara langsung mengambil bagian dalam tugas akhir ini:

- Bapak Prof. Arif Djunaidy selaku dosen pembimbing pertama dan Ibu Renny Pradina Kusumawardani selaku dosen pembimbing kedua yang telah sudi meluangkan waktu untuk membantu dan memberikan bimbingan, ilmu, nasehat, serta motivasi kepada penulis.
- Ibu Lilik Malikhah selaku ibunda penulis dan Ayah Mochamad Ali selaku ayahanda penulis yang telah memberikan doa, doa, dan doa serta bantuan secara materiil yang sangat berguna sehingga penulis dapat menyelesaikan pendidikan khususnya Tugas Akhir ini.
- Mas Agil, Mas Kahfi, Edo, dan Arsyal selaku saudara kandung alias saudara *pendowo limo* penulis yang senantiasa mendukung penulis kapan pun.

- Ibu Mahendrawathi Erawan, Bapak Rully Agus Hendrawan dan Ibu Retno Aulia Vinarti selaku dosen penguji yang telah memberikan kritik dan saran untuk perbaikan Tugas Akhir ini.
- Ibu Feby Artwodini Muqtadiroh selaku dosen wali penulis yang sudah banyak memberikan masukan selama proses awal perkuliahan hingga pengerjaan Tugas Akhir ini.
- Teman-teman sebidang minat SPK-IB, Mamed, Miya, Shinta, Icha, Ejak, Eka, Alin, Izmi, Ela, Nia, Amal, Pipik, Muci, Budhe, Ikar, Suvi, dan Mas Juhdi, yang sudah memberikan dukungan secara lahir dalam menyelesaikan Tugas Akhir ini.
- Teman-teman dari bidang minat PPSI dan EBIS, Fery, Bisma, Ebik, Faridl, Adhika, Yoga, Rijal, Om, Adis, Opik, Syum, Yan, Yance, Fanyo, dan Djay, yang sudah memberikan semangat dan dukungan dalam pengerjaan Tugas Akhir ini.
- Sahabat seperjuangan, Vinda dan Amel, yang sudah bersama-sama berjibaku dan memberikan semangat dalam mengerjakan Tugas Akhir sekaligus perkuliahan di Program Pascasarjana Magister Jalur Cepat ITS.
- Saudara pengajian *Nashoihud Diniyah*, Hamim, Abdan, dan Danis, yang sudah mengingatkan dan memberikan dukungan secara batin selama pengerjaan Tugas Akhir ini.
- Serta seluruh pihak yang belum penulis sebutkan satu persatu.

Terima kasih atas segala bantuan, dukungan, serta doanya. Semoga Allah SWT. senantiasa melimpahkan rahmat hidayah serta membalas kebaikan-kebaikan yang telah diberikan kepada penulis.

Surabaya, 18 Juli 2014

Penulis

## DAFTAR ISI

Abstrak .....	vii
Abstract .....	ix
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
DAFTAR GAMBAR .....	xvi
DAFTAR TABEL .....	xviii
DAFTAR KODE.....	xix
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1. Latar belakang.....	1
1.2. Perumusan Masalah .....	5
1.3. Batasan Tugas Akhir.....	5
1.4. Tujuan Tugas Akhir .....	5
1.5. Relevansi atau Manfaat Tugas Akhir.....	5
1.6. Sistematika Penulisan .....	6
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>9</b>
2.1. Twitter API Stream.....	9
2.2. Analisis Sentimen .....	11
2.3. Klasifikasi Teks .....	12
2.3.1. Metode <i>Naïve Bayes</i> .....	12
2.3.2. Evaluasi Metoda.....	18
2.4. Aplikasi OpenGeo Suite .....	21
2.5. <i>Heat Map</i> .....	24
2.6. <i>Customer Relationship Management (CRM)</i> .....	26
2.7. Operasional CRM PT. Telkomsel .....	27
<b>BAB III METODOLOGI Pengerjaan Tugas Akhir.....</b>	<b>29</b>
3.1. Studi Literatur .....	30
3.2. Pengumpulan Data.....	30
3.2.1. Persiapan Twitter API <i>Stream</i> .....	31
3.2.2. Proses Twitter API <i>Stream</i> .....	33
3.3. Praproses Data .....	35
3.3.1. Pemilihan Atribut Data.....	35
3.3.2. Eliminasi Data tanpa Koordinat .....	35

3.3.3.	Pemilahan Data Opini dan Non-Opini .....	36
3.3.4.	Pemilihan Data Opini dengan Aspek Operasional CRM.....	37
3.4.	Klasifikasi Teks.....	37
3.5.	Pembuatan <i>Heat Map</i> .....	38
3.6.	Pembuatan Laporan.....	39
<b>BAB IV</b>	<b>DESAIN SISTEM.....</b>	<b>41</b>
4.1.	Desain Aplikasi Pengklasifikasi Teks .....	41
4.1.1.	Desain Basis Data Pengklasifikasi Teks.....	41
4.1.2.	Diagram <i>Use Case</i> Pengklasifikasi Teks.....	42
4.1.3.	Diagram Sekuen Pengklasifikasi Teks .....	43
4.1.4.	Diagram Aktivitas Pengklasifikasi Teks .....	46
4.2.	Desain Halaman Web <i>Heat Map</i> .....	48
4.2.1.	Desain Basis Data Halaman Web <i>Heat Map</i> .....	48
4.2.2.	Diagram Use Case Halaman Web <i>Heat Map</i> .....	49
4.2.3.	Diagram Sekuen Melihat Pemetaan Data.....	50
4.2.4.	Diagram Aktivitas Melihat Pemetaan Data.....	51
<b>BAB V</b>	<b>IMPLEMENTASI APLIKASI.....</b>	<b>53</b>
5.1.	Implementasi Aplikasi Pengklasifikasi Teks .....	53
5.1.1.	Pembuatan Basis Data di MySQL.....	53
5.1.2.	Implementasi Fungsi Mengunggah Data Dokumen .....	57
5.1.3.	Implementasi Pengklasifikasi Teks .....	58
5.2.	Implementasi Halaman Web <i>Heat Map</i> .....	68
5.2.1.	Pembuatan Basis Data di PostGIS.....	68
5.2.2.	Penggabungan Atribut Koordinat.....	70
5.2.3.	Pengaturan <i>Heat Map</i> di GeoServer.....	72
5.2.4.	Penyusunan Lapisan di GeoExplorer .....	77
5.2.5.	Pembuatan <i>Web Page</i> untuk Pemetaan Data.....	78
<b>BAB VI</b>	<b>UJI COBA DAN ANALISIS HASIL .....</b>	<b>85</b>
6.1.	Data Uji Coba.....	85
6.2.	Lingkungan Uji Coba.....	87
6.3.	Skenario Uji Coba.....	88
6.3.1.	Uji Validasi Aplikasi Pengklasifikasi Teks .....	88
6.3.2.	Uji Validasi Pemetaan dalam Bentuk <i>Heat Map</i> ..	88

6.4.	Pelaksanaan dan Hasil.....	89
6.4.1.	Hasil Uji Coba Aplikasi Pengklasifikasi Teks .....	89
6.4.2.	Hasil Uji Coba Pemetaan dalam Bentuk <i>Heat Map</i> .....	91
6.5.	Analisis Hasil.....	93
6.5.1.	Analisis Aplikasi Pengklasifikasi Teks .....	93
6.5.2.	Analisis Pemetaan dalam Bentuk <i>Heat Map</i> .....	97
BAB VII KESIMPULAN .....		101
DAFTAR PUSTAKA.....		103
RIWAYAT PENULIS.....		107
LAMPIRAN A - DATA FITUR PELATIHAN.....		A-1

## DAFTAR TABEL

Tabel 2.1. Hasil Proses Pembelajaran (1).....	16
Tabel 2.2. Hasil Proses Pembelajaran (2).....	17
Tabel 2.3. Hasil Proses Pembelajaran (3).....	17
Tabel 2.4. Klasifikasi Dokumen .....	19
Tabel 2.5. Tabel Kontingensi .....	20
Tabel 3.1. Pemilahan Data Opini.....	36
Tabel 4.1. Desain Basis Data untuk Tabel 'tweet'.....	42
Tabel 4.2. Desain Basis Data untuk Tabel 'keyword' .....	42
Tabel 4.3. Desain Basis Data untuk Halaman Web Heat Map ....	48
Tabel 4.4. Desain Basis Data untuk Pemetaan Data.....	49
Tabel 5.1. Kategori Fitur Operasional CRM .....	54
Tabel 5.2. Basis Data Fitur Kata.....	55
Tabel 5.3. Basis Data Fitur <i>Emoticons</i> .....	56
Tabel 5.4. Data Contoh Sebelum diklasifikasi .....	67
Tabel 5.5. Data Contoh Hasil Klasifikasi .....	67
Tabel 5.6. Hasil Klasifikasi di dalam Basis Data .....	68
Tabel 6.1. Rincian Jumlah Data Latih dan Data Uji Coba .....	86
Tabel 6.2. Nilai Jarak untuk Pembuatan Dendogram .....	86
Tabel 6.3. Perangkat Keras Lingkungan Uji Coba.....	87
Tabel 6.4. Perangkat Lunak Lingkungan Uji Coba.....	87
Tabel 6.5. Tabel Kontingensi Hasil Klasifikasi Teks .....	89
Tabel 6.6. Contoh Tweet dengan Bahasa Tidak Baku.....	94
Tabel 6.7. Contoh Tweet Fenomena Sarkasme .....	95
Tabel A-1. Fitur Pilihan Untuk Data Pelatihan .....	A-1

## DAFTAR GAMBAR

Gambar 1.1. Grafik Pengguna Twitter (SemioCast, 2012) .....	2
Gambar 1.2. Grafik <i>Tweets Posted</i> (SemioCast, 2012).....	3
Gambar 2.1. Mekanisme Twitter API <i>Stream</i> (dev.twitter.com)..	10
Gambar 2.2. Proses Pembelajaran <i>Naïve Bayes</i> (Lestari, Putra, & Chayawan, 2013).....	14
Gambar 2.3. Proses Klasifikasi <i>Naïve Bayes</i> (Lestari, Putra, & Chayawan, 2013).....	15
Gambar 2.4. Arsitektur OpenGeo Suite (BoundlessGeo, 2014)..	22
Gambar 2.5. Fitur-fitur OpenGeo Suite (BoundlessGeo, 2014)..	23
Gambar 2.6. Tampilan <i>Heat Map</i> .....	23
Gambar 2.7. Contoh Interpretasi Warna.....	25
Gambar 2.8. Dendrogram <i>Hierarchical Clustering</i> .....	26
Gambar 2.9. Logo PT. Telkomsel .....	27
Gambar 2.10. Logo Telkomsel 2.0 .....	28
Gambar 3.1. Metodologi Pengerjaan Tugas Akhir.....	29
Gambar 3.2. Nomor <i>Token</i> Twitter API.....	31
Gambar 3.3. Ilustrasi Proses Pengumpulan Data Twitter.....	34
Gambar 4.1. Diagram <i>Use Case</i> Pengklasifikasi Teks .....	43
Gambar 4.2. Diagram Sekuen Mengunggah Data Dokumen .....	44
Gambar 4.3. Diagram Sekuen Melakukan Klasifikasi Teks.....	45
Gambar 4.4. Diagram Aktivitas Mengunggah Data Dokumen ...	46
Gambar 4.5. Diagram Aktivitas Melakukan Klasifikasi Teks.....	47
Gambar 4.6. Diagram <i>Use Case</i> Halaman Web <i>Heat Map</i> .....	50
Gambar 4.7. Diagram Sekuen Melihat Pemetaan Data .....	50
Gambar 4.8. Diagram Aktivitas Melihat Pemetaan Data .....	51
Gambar 5.3. Program Klasifikasi Teks <i>Naive Bayes</i> .....	66
Gambar 5.2. Data Hasil Impor ke PostGIS .....	70
Gambar 5.3. Data Hasil Penggabungan Koordinat.....	71
Gambar 5.4. Klaster Sentimen Negatif.....	74
Gambar 5.5. Klaster Sentimen Positif .....	74
Gambar 5.6. Titik-titik Sentimen Negatif.....	75
Gambar 5.7. Titik-titik Sentimen Positif .....	75

Gambar 5.8. Peta OpenStreetMap .....	76
Gambar 5.9. Peta Bing Aerial with Labels .....	76
Gambar 5.10. Lapisan Pemetaan .....	77
Gambar 5.11. Tampilan Pemetaan Basis Peta OpenStreetMap... 78	
Gambar 5.12. Tampilan Pemetaan Basis Peta Bing Aerial with Labels .....	78
Gambar 5.13. Tampilan Pemetaan dalam GeoExplorer .....	79
Gambar 5.14. Fitur <i>Publish Map</i> GeoExplorer .....	80
Gambar 5.15. Link Peta dari GeoExplorer .....	80
Gambar 5.16. Tampilan Pemetaan dalam <i>Web Page</i> .....	84
Gambar 6.1. Dendogram <i>Heat Map</i> yang dihasilkan OpenGeo Suite.....	91
Gambar 6.2. Klaster <i>Heat Map</i> dalam Skala 1.000 m.....	92
Gambar 6.3. Klaster <i>Heat Map</i> dalam Skala 2.000 m.....	92
Gambar 6.4. Klaster <i>Heat Map</i> dalam Skala 5.000 m.....	93
Gambar 6.5. Pemetaan Data Sentimen Positif di Periode Desember.....	98
Gambar 6.6. Pemetaan Data Sentimen Positif di Periode April..	98
Gambar 6.7. Pemetaan Data Sentimen Negatif di Periode Desember.....	99
Gambar 6.8. Pemetaan Data Sentimen Negatif di Periode April	99



## DAFTAR KODE

Kode 5.1. Query untuk Mengunggah Data Dokumen .....	57
Kode 5.2. <i>Query</i> Metode Naïve Bayes .....	59
Kode 5.3. <i>Query</i> Inisiasi Variabel Naïve Bayes .....	60
Kode 5.4. <i>Query</i> Fungsi <i>get_sentiment_result</i> .....	61
Kode 5.5. <i>Query</i> Perhitungan Fitur Naïve Bayes .....	61
Kode 5.6. <i>Query</i> Hasil Metode Naïve Bayes .....	62
Kode 5.7. <i>Query text_model.php</i> .....	62
Kode 5.8. <i>Query</i> untuk Koneksi ke Basis Data .....	64
Kode 5.9. <i>Query</i> Fungsi <i>get_all_keywords</i> .....	64
Kode 5.10. <i>Query</i> untuk Total Frekuensi Fitur .....	65
Kode 5.11. <i>Query</i> Fungsi <i>get_total_sum</i> .....	65
Kode 5.12. <i>Query membuat_tabel.sql</i> .....	69
Kode 5.13. <i>Query impor_data.sql</i> .....	69
Kode 5.14. <i>Query membuat_tabel_koordinat.sql</i> .....	70
Kode 5.15. <i>Query penggabungan_koordinat.sql</i> .....	71
Kode 5.16. <i>Query</i> Pemberian Warna <i>Heat Map</i> .....	73
Kode 5.17. <i>Query</i> Pengaturan Radius <i>Heat Map</i> .....	73
Kode 5.18. <i>Query</i> Link Peta dari GeoExplorer .....	81
Kode 5.19. <i>Query lihat_peta.php</i> .....	81

# **BAB I**

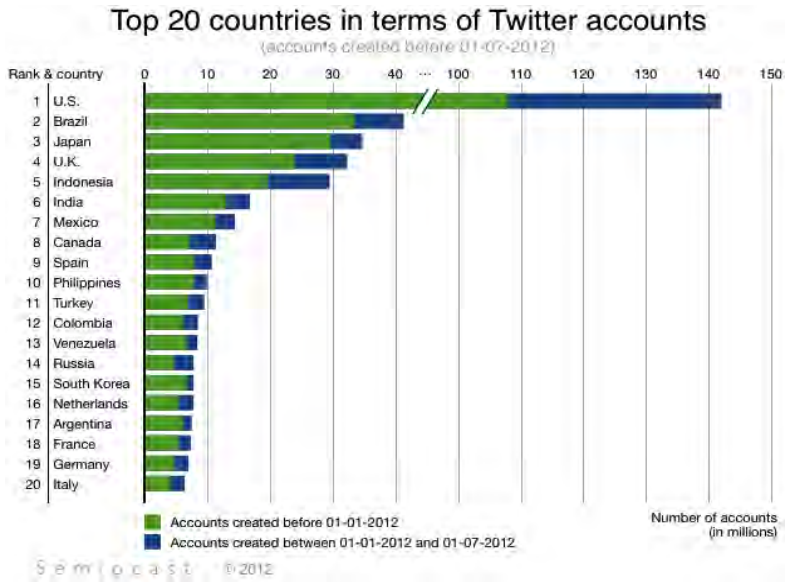
## **PENDAHULUAN**

Bab pendahuluan ini membahas mengenai latar belakang, perumusan masalah, batasan tugas akhir, tujuan tugas akhir, relevansi atau manfaat dari pengerjaan tugas akhir, serta sistematika penulisan.

### **1.1. Latar belakang**

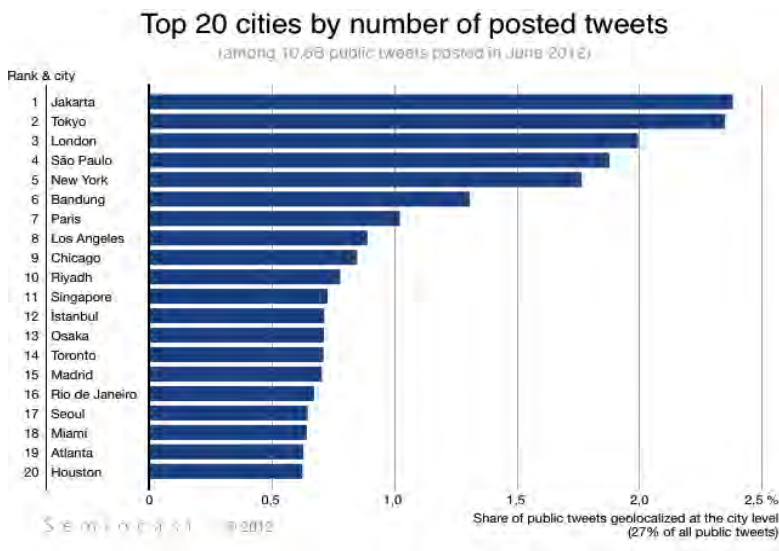
Pentingnya *Customer Relationship Management* (CRM) atau pengelolaan hubungan pelanggan bagi perusahaan adalah sebuah strategi bisnis menyeluruh dalam suatu perusahaan yang memungkinkan perusahaan secara efektif mengelola dan menjaga hubungan dengan pelanggan. Dengan CRM, perusahaan dapat mengetahui apa saja kekurangan dari layanan dan perbaikan apa yang harus dilakukan. Terciptanya loyalitas pelanggan disebabkan karena meningkatnya kepuasan pelanggan atau beralihnya pelanggan tidak puas menjadi puas atas pelayanan yang diberikan perusahaan. CRM ini dapat dilakukan dengan mengumpulkan segala bentuk interaksi pelanggan baik itu lewat telepon, e-mail, maupun masukan di media jejaring sosial (Yahya, 2013).

Twitter merupakan salah satu jejaring sosial yang mendunia. Hampir dari seluruh masyarakat di berbagai negara menggunakannya. Berdasarkan hasil perhitungan tahun 2012 oleh SemioCast, perusahaan peneliti media sosial di Prancis, Indonesia termasuk negara pengguna terbanyak ke lima di dunia (lihat gambar 1.1). Twitter oleh sebagian masyarakat Indonesia dijadikan sebagai media untuk menyampaikan opini. Opini dapat digolongkan menjadi opini positif berupa pujian dan opini negatif berupa kritik. Di dalam Twitter, banyak opini positif dan negatif yang dilayangkan pada perusahaan atas kualitas layanannya. Opini masyarakat ini sekaligus menjadi bahan evaluasi bagi perusahaan untuk perbaikan dan peningkatan kualitas layanan ke depannya.



**Gambar 1.1. Grafik Pengguna Twitter  
(Semio cast, 2012)**

Apabila opini-opini di dalam Twitter dikumpulkan, maka opini tersebut akan menjadi himpunan data besar yang dapat mewakili pandangan mayoritas pelanggan terhadap layanan perusahaan. Di Indonesia, ada dua kota yang termasuk dalam 20 besar kota yang paling banyak melayangkan *tweet*. Gambar 1.2 menunjukkan Kota Jakarta adalah kota yang terbanyak melayangkan *tweet*, sedangkan Kota Bandung berada pada urutan ke-6. Hal ini dapat merepresentasikan bahwa pengguna Twitter di Indonesia relatif banyak. Oleh karena itu, pengumpulan data melalui Twitter menjadi proses yang bernilai untuk diambil manfaatnya dalam konteks pengelolaan hubungan pelanggan. Selain itu, di dalam Twitter juga ada fitur lokasi *tweet*. Dengan fitur ini, posisi pelanggan dapat diketahui dengan melihat koordinat lintang dan bujur sehingga perusahaan dapat mengetahui lokasi persis orang yang melayangkan *tweet*.



**Gambar 1.2. Grafik Tweets Posted**  
(SemioCast, 2012)

Oleh karena manfaat tersebut di atas, proses dalam menerapkannya diperlukan cara tertentu yang mampu merealisasikan hal tersebut. Semakin banyak orang yang berpendapat, maka semakin representatif data yang diperoleh. Berdasarkan permasalahan-permasalahan yang telah dipaparkan di awal, proses dalam menerapkannya membutuhkan cara yang mampu mengumpulkan data Twitter. Hal ini dapat dilakukan dengan memanfaatkan Twitter API *Stream*. Selain itu, proses untuk mengklasifikasikan opini ke dalam klasifikasi opini positif atau opini negatif juga diperlukan. Untuk hal ini, metode dalam mengklasifikasikan opini atau biasa disebut klasifikasi teks menggunakan *Naïve Bayes*. Terakhir, proses yang memanfaatkan fitur lokasi *tweet* dapat dilakukan dengan pemetaan data Twitter dalam bentuk *heat map*. Penerapannya dapat memanfaatkan aplikasi OpenGeo Suite. Aplikasi *open source* ini memungkinkan data Twitter yang sudah diklasifikasi untuk ditampilkan ke dalam *heat map*.

Selain *Naïve Bayes* juga ada metode-metode lain yang pernah digunakan untuk melakukan klasifikasi teks seperti *n-gram analysis* dan *ontology-based*. Metode *n-gram analysis* adalah pendekatan klasifikasi teks yang digunakan untuk mengekstrak atau mengambil ciri khusus dari suatu obyek (Ghiassi, Skinner, & Zimbra, 2013). Sedangkan *ontology-based* berfokus pada pemodelan istilah-istilah ke dalam suatu domain (Kontopoulos, Berberidis, Dergiades, & Bassiliades, 2013). Namun selama ini kedua metode tersebut hanya digunakan untuk teks berbahasa Inggris atau bahasa lainnya selain bahasa Indonesia. Untuk itu, dalam persoalan ini, metode yang akan digunakan adalah *Naïve Bayes*. Calvin, Setiawan, dan Johan (2014) menunjukkan bahwa metode ini berfungsi baik untuk mengklasifikasikan obyek berbentuk teks bahasa Indonesia di Twitter. Sehingga metode ini memungkinkan untuk menyelesaikan persoalan yang telah dipaparkan sebelumnya.

*Heat map* adalah salah satu bentuk visualisasi yang menampilkan gradasi warna dari pengelompokan data berdasarkan faktor tertentu. Pengelompokan atau *clustering* dalam *heat map* dapat berdasarkan dari faktor tingkat kepadatan data. Perpaduan pengelompokan data dengan gradasi warna ini memudahkan dalam melihat tingkat kepadatan pengguna Twitter yang melayangkan *tweet* pada sebuah peta wilayah. Oleh karena itu pemetaan data Twitter akan disajikan dalam bentuk *heat map*.

Penerapan klasifikasi teks untuk pemetaan data Twitter pelanggan dalam bentuk *heat map* ini diharapkan dapat membantu Telkomsel dalam memperoleh manfaat dari CRM. Manfaat yang diharapkan seperti membantu Telkomsel untuk mengetahui perubahan tingkat kepuasan dan ketidakpuasan pelanggan pada periode-periode tertentu, untuk meningkatkan usaha pemasaran (*marketing effort*), atau untuk mengetahui wilayah dengan mayoritas pelanggan mengeluh atas layanan yang mungkin disebabkan oleh berbagai faktor seperti infrastruktur kurang memadai, permintaan melebihi kapasitas, dsb.

## 1.2. Perumusan Masalah

Berdasarkan paparan di latar belakang, perumusan masalah dalam pengerjaan tugas akhir ini adalah sebagai berikut:

- Bagaimana melakukan praproses data dari jejaring sosial Twitter agar siap diproses oleh aplikasi pengklasifikasi teks dan dapat dipetakan ke dalam *heat map*.
- Bagaimana menampilkan hasil pemetaan *heat map* dalam bentuk *web page*.

## 1.3. Batasan Tugas Akhir

Batasan yang ada dalam pengerjaan tugas akhir ini adalah sebagai berikut:

- Data yang dikumpulkan bersumber dari Twitter.
- Peta yang ditampilkan meliputi wilayah-wilayah di Indonesia.
- Pengambilan kata kunci dalam metode *Naive Bayes* didasarkan justifikasi terkait dengan layanan operasional.

## 1.4. Tujuan Tugas Akhir

Tujuan dari tugas akhir ini adalah untuk menghasilkan visualisasi pemetaan data opini pelanggan Telkomsel di Twitter dalam bentuk *heat map* yang sudah diklasifikasi ke dalam kelas opini positif dan opini negatif.

## 1.5. Relevansi atau Manfaat Tugas Akhir

Dengan memetakan data Twitter pelanggan Telkomsel dalam bentuk *heat map* ini diharapkan dapat memudahkan Telkomsel dalam menerapkan operasional CRM yakni untuk mengetahui lokasi-lokasi pelanggan yang puas dan tidak puas terhadap layanannya. Tantangan yang dihadapi Telkomsel untuk mencapai tujuan strategis dalam rangka mengembangkan dan mematangkan segmen pasar adalah tingginya pertumbuhan ekspektasi kepuasan pelanggan dan semakin canggihnya alat komunikasi (*device*) untuk mengakses berbagai layanan Telkomsel (PT. Telekomunikasi Selular *Annual Report*, 2012).

Hasil tugas akhir ini berkenaan dengan pengelolaan hubungan dengan pelanggan terkait dengan letak geografisnya. Sehingga diharapkan tugas akhir ini dapat membantu Telkomsel dalam mencapai tujuan strategisnya.

## **1.6. Sistematika Penulisan**

Sistematika penulisan tugas akhir ini dibagi menjadi enam bab sebagai berikut:

### **BAB I PENDAHULUAN**

Bab ini menjelaskan mengenai latar belakang, perumusan masalah, batasan tugas akhir, tujuan tugas akhir, relevansi atau manfaat tugas akhir bagi Telkomsel, dan sistematika penulisan yang diterapkan dalam pemaparan tugas akhir.

### **BAB II TINJAUAN PUSTAKA**

Bab ini menjelaskan mengenai peninjauan literatur dan uraian penjelasan yang digunakan dalam pengerjaan tugas akhir ini yang meliputi Twitter API *Stream*, analisis sentimen, klasifikasi teks, aplikasi OpenGeo Suite, *heat map*, CRM, dan operasional CRM PT. Telkomsel.

### **BAB III METODOLOGI PENELITIAN**

Bab ini menjelaskan mengenai metodologi pengerjaan tugas akhir. Metodologi tugas akhir terdiri dari beberapa langkah pengerjaan tugas akhir yaitu studi literatur, pengumpulan data, praproses data, klasifikasi teks, dan pembuatan *heat map*

### **BAB IV DESAIN SISTEM**

Bab ini menjelaskan mengenai desain sistem dari aplikasi klasifikasi teks berbasis *web* dengan menggunakan metode *Naïve Bayes*. Desain ini yang kemudian diimplementasikan pada bab implementasi aplikasi

## **BAB V IMPLEMENTASI PERANGKAT LUNAK**

Bab ini menjelaskan mengenai implementasi perangkat lunak klasifikasi teks, pemetaan hasil klasifikasi teks menggunakan aplikasi OpenGeo Suite, dan pembuatan *web page* hasil pemetaan.

## **BAB VI UJI COBA DAN ANALISIS HASIL**

Bab ini menjelaskan mengenai uji validasi pada aplikasi klasifikasi teks. Ada tiga uji coba yang dilakukan yaitu uji akurasi, uji presisi, dan uji *recall*. Selain itu juga dilakukan analisis hasil dari keseluruhan pengujian.

## **BAB VII KESIMPULAN DAN SARAN**

Bab ini menjelaskan mengenai kesimpulan yang dihasilkan dari pengujian aplikasi klasifikasi teks beserta saran untuk perbaikan selanjutnya.



*Halaman ini sengaja dikosongkan*

## **BAB II**

### **TINJAUAN PUSTAKA**

Bab ini membahas mengenai hasil studi literatur yang digunakan dalam pengerjaan tugas akhir yang meliputi Twitter API *Stream*, analisis sentimen, klasifikasi teks, aplikasi OpenGeo Suite, *Heat Map*, CRM, dan operasional CRM PT. Telkomsel.

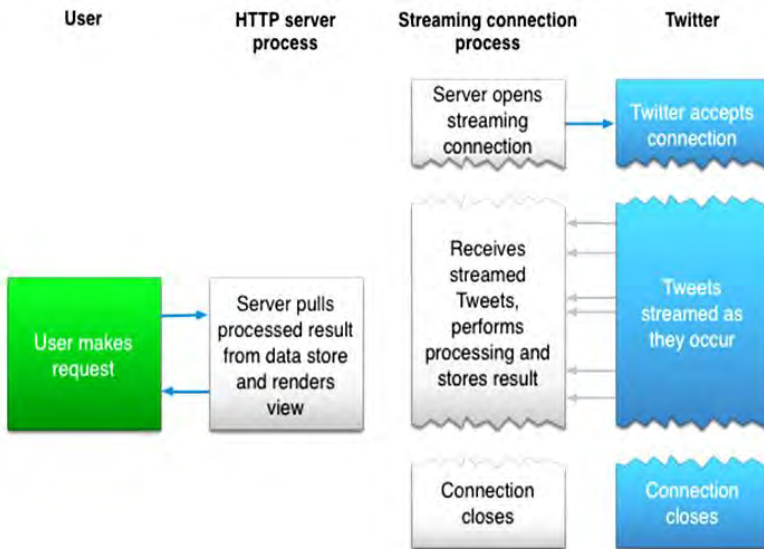
#### **2.1. Twitter API Stream**

Twitter API *Stream* adalah sebuah program untuk mendapatkan data Twitter dengan meminta akses layanan menggunakan *Application Programming Interfaces (API)*. Twitter, jejaring sosial yang berbasis layanan situs (*web services*), menggunakan API untuk membolehkan aplikasi lain menggunakan atau berinteraksi dengan datanya (*third-party applications*). Matthew C. Wagner (2013) menyatakan bahwa API dari Twitter berperan sebagai jembatan yang memfasilitasi interaksi antara Twitter dengan program lain. Di dalam Twitter, representasi API berupa nomor-nomor unik (*token*) yang hanya dapat digunakan oleh sebuah program saja. Twitter API *Stream* ini bekerja dalam notasi komputasi *python*. Bahasa pemrograman *python* ini mampu mengakomodasi terkait pengumpulan data dalam jumlah besar dengan memanfaatkan Twitter API.

Untuk mengambil data dari Twitter, setiap pengembang perangkat lunak harus mengikuti aturan yang telah ditetapkan pengembang Twitter, termasuk penerapan Twitter API *Stream* ini. Karena API itu sendiri terdiri dari sebuah himpunan peraturan dari seorang pengembang perangkat lunak yang harus diikuti oleh pengembang perangkat lunak lain yang ingin berinteraksi dengan datanya (Wagner, 2013). Versi Twitter API yang digunakan harus yang paling baru. Saat ini, versi Twitter API yang terbaru adalah versi 1.1. Selain itu, data yang diperbolehkan untuk

diambil adalah data waktu nyata (*real time*) sehingga data historis atau data yang sudah lampau tidak dapat diambil kembali.

Pada dasarnya, proses pengambilan data di mulai dari pengguna membuat permintaan kepada HTTP *server* hingga kemudian HTTP *server* mengembalikan hasil dari pengambilan data kepada pengguna. Seperti ditunjukkan gambar 2.1 proses pengambilan data di mulai dengan membuat koneksi ke Twitter hingga Twitter menerima koneksi tersebut. Kemudian data *tweet* dapat diterima untuk kemudian dapat langsung diproses oleh aplikasi atau sekadar disimpan. Proses ini terus berlangsung selama masih ada koneksi baik koneksi jaringan internet maupun koneksi Twitter API. Pengambilan data baru akan terhenti jika salah satu dari koneksi tersebut terputus.



Gambar 2.1. Mekanisme Twitter API *Stream*  
(dev.twitter.com)

## 2.2. Analisis Sentimen

Analisis sentimen (*sentiment analysis*), atau biasa disebut dengan *opinion mining*, adalah sebuah bidang studi yang mana menganalisis opini, sentimen, evaluasi, pujian, emosi, dan sikap orang terhadap sesuatu seperti produk, layanan, organisasi, individu, isu, acara, topik, dan lain-lain. Klasifikasi teks utamanya berfokus pada opini yang mana mengekspresikan atau menyimpulkan sentimen positif atau negatif. Dengan berkembang pesatnya aplikasi-aplikasi sosial media seperti Friendster, Blog, Facebook, Twitter, dll maka semakin besar pula kebutuhan organisasi untuk memanfaatkan teknologi ini (Bing Liu, 2012). Secara umum analisis sentimen dibedakan menjadi tiga tingkatan, antara lain:

- **Tingkatan Dokumen.** Tingkatan ini mengklasifikasikan apakah sebuah dokumen mengekspresikan sentimen positif atau negatif. Misalkan diberikan sebuah tinjauan produk, sistem menentukan apakah tinjauan tersebut mengekspresikan secara keseluruhan merupakan opini positif atau negatif tentang produk itu. Sehingga pada tingkatan ini suatu dokumen diklasifikasikan menjadi dua klasifikasi. Tingkatan ini disebut juga klasifikasi sentimen tingkatan dokumen
- **Tingkatan Kalimat.** Tingkatan ini melihat suatu dokumen, lalu menentukan apakah tiap dokumen mengekspresikan opini positif, negatif, atau netral. Netral biasanya berarti tidak berpendapat. Analisis tingkatan ini erat kaitannya dengan klasifikasi subyektifitas (*subjectivity classification*).
- **Tingkatan Entitas dan Aspek.** Tingkatan yang hasil analisis opininya tidak ditemukan pada tingkatan dokumen dan tingkatan kalimat. Aspek ini memunculkan implikasi lain, selain sentimen positif atau negatif, berupa target dari opini. Tingkatan ini memperhatikan konstruksi bahasa lebih mendetail, seperti klausa, dan frase kalimat. Tingkatan ini disebut juga dengan tingkatan fitur (*feature level*) atau fitur berbasis penggalan opini dan ringkasan (*feature-based opinion mining and summarization*).

### 2.3. Klasifikasi Teks

Klasifikasi teks (*text classification*) atau juga dikenal dengan kategorisasi teks adalah bagian dari pengelompokan dokumen atau teks dengan memperhatikan kategori. Secara umum, klasifikasi teks termasuk penggolongan teks berbasis topik dan penggolongan teks berbasis kategori (*genre*) (Ikonomakis, Kotsiantis, & Tampakas, 2005). Teks dapat dikategorikan menjadi artikel ilmiah, laporan surat kabar, tinjauan film, iklan, dll. Klasifikasi teks dapat menggunakan banyak teknik pembelajaran mesin. Pendekatan yang sering diadopsi adalah pohon keputusan (*decision tree*), induksi aturan (*rule induction*), jaringan saraf (*neural network*), *Support Vector Machines*, dan *Naïve Bayes*.

#### 2.3.1. Metode *Naïve Bayes*

*Naïve Bayes* adalah metode yang paling sering digunakan dalam menerapkan klasifikasi teks karena kesederhanaan dan keefektifan metodenya (Ikonomakis, Kotsiantis, & Tampakas, 2005). Nama metode ini diambil dari nama penemu metode tersebut pada tahun 1702 yaitu Thomas Bayes. Dalam penerapannya, metode *Naïve Bayes* selain untuk mengklasifikasikan teks dalam konteks analisis sentimen, metode ini juga digunakan untuk melakukan penyaringan email, kategorisasi subyek, identifikasi penulis, identifikasi jenis kelamin atau usia, dan identifikasi bahasa.

Metode pembelajaran mesin ini dilakukan berdasarkan data pelatihan dengan menggunakan kemungkinan kondisional sebagai dasar. Metode ini bisa melakukan klasifikasi atau kategorisasi dari beberapa kata teks yang menghasilkan sama seperti dalam tingkatan dokumen. Metode *Naïve Bayes* terdiri dari dua fase yaitu fase pembelajaran dan fase klasifikasi. Fase pembelajaran adalah fase dimana dokumen hasil praproses melewati proses pembelajaran untuk mendapatkan data pelatihan. Sedangkan fase klasifikasi adalah fase dimana data baru akan

diklasifikasikan menggunakan data pelatihan yang telah dibuat sebelumnya (Lestari, Putra, & Chayawan, 2013).

Fase pembelajaran dilakukan setelah dokumen melalui data praproses. Proses fase pembelajaran ini dilakukan secara manual. Awalnya sebuah sampel dokumen, yang mana semakin banyak jumlahnya semakin baik, ditentukan masing-masing klasifikasinya. Kata-kata unik yang terdapat di dalam dokumen dimasukkan ke dalam basis data sebagai klasifikasi positif atau negatif. Data pelatihan inilah yang akan digunakan untuk mengolah dokumen baru pada fase klasifikasi. Diagram alur proses fase pembelajaran dapat dilihat pada gambar 2.2, sedangkan formulanya lihat persamaan 1.

$$P(V_j) = \frac{|fd(V_j)|}{|D|} \quad (1)$$

Persamaan  $P(V_j)$  digunakan untuk menghitung perbandingan dari jumlah kelas dokumen yang ada. Persamaan 1 secara berturut-turut didefinisikan  $|fd(V_j)|$ , sebuah kelas dokumen, dan  $|D|$ , jumlah kelas dokumen untuk dijadikan model. Kemudian persamaan yang digunakan untuk menghitung indeks kelas dengan melibatkan frekuensi kata kunci dokumen dituliskan dengan  $P(W_k|V_j)$  seperti terlihat persamaan 2. Persamaan  $P(W_k|V_j)$  ini adalah hasil dari  $f(W_k|V_j)$ , jumlah frekuensi kemunculan kata unik pada kelas tertentu dengan ditambahkan 1, dibagi dengan jumlah dari  $N$ , jumlah frekuensi kata unik dalam data pelatihan, dan  $|W|$ , jumlah frekuensi kata unik setiap ada penambahan yang muncul. Untuk kedua persamaan tersebut, implementasinya membutuhkan basis data yang menampung frekuensi kata unik tersebut pada tiap-tiap dokumen.

$$P(W_k|V_j) = \frac{f(W_k|V_j) + 1}{N + |W|} \quad (2)$$

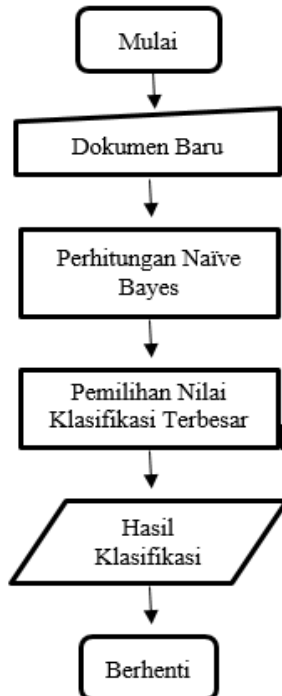


**Gambar 2.2.** Proses Pembelajaran *Naïve Bayes*  
(Lestari, Putra, & Chayawan, 2013)

Berdasarkan kedua persamaan tersebut, klasifikasi teks menghasilkan keluaran berupa sentimen positif atau negatif terhadap sebuah dokumen. Fase pembelajaran ini menghasilkan data pelatihan yang memiliki model probabilistik sesuai dengan dokumen sampel yang telah dimasukkan di awal. Dokumen yang baru akan melalui proses fase klasifikasi dengan menggunakan data pelatihan (Lestari, Putra, & Chayawan, 2013). Kecukupan data pelatihan menjadi penting untuk meningkatkan akurasi mesin pembelajaran. Semakin banyak data pelatihan, maka semakin

tinggi akurasi mesin pembelajaran *Naïve Bayes*. Ini juga diperlukan untuk menyelesaikan persoalan klasifikasi teks berbahasa jejaring sosial yang diketahui banyak kata-kata tidak baku di dalamnya.

Pada fase klasifikasi, dokumen juga melalui praproses data terlebih dahulu. Proses ini dilakukan berdasarkan data pelatihan yang telah dihasilkan oleh fase pembelajaran. Proses dimulai dengan memasukkan dokumen baru ke perhitungan *Naïve Bayes* untuk dihitung nilai tiap-tiap klasifikasinya. Kemudian hasil perhitungan dengan nilai terbesar dipilih sebagai hasil klasifikasi. Diagram alur fase klasifikasi dapat dilihat gambar 2.3.



Gambar 2.3. Proses Klasifikasi *Naïve Bayes*  
(Lestari, Putra, & Chayawan, 2013)



Di dalam literatur Lestari, Putra, & Cahyawan (2013), terdapat contoh dokumen pengujian berupa kepribadian seseorang yang memiliki klasifikasi antara sanguinis, koleris, melankolis, dan plagmatis. Namun data dokumen tersebut belum diketahui klasifikasinya. Maka Lestari dkk mengklasifikasikan data dokumen tersebut dengan menggunakan metode *Naïve Bayes*.

Sebagai permisalan, menurut Lestari, Putra, & Cahyawan (2013), ada sebuah masukan dokumen data contoh kepribadian yang belum diketahui klasifikasinya sebagai berikut: “*Saya adalah orang yang jujur, ceria, ramah, sabar, dan humoris. Saya suka bergaul dengan teman. Saya suka bertualang. Saya suka mengenal hal baru tetapi saya juga sering bersedih.*” Dokumen ini kemudian diklasifikasi menggunakan metode *Naïve Bayes* seperti ditunjukkan tabel tabel 2.1, 2.2, dan 2.3.

Apabila sudah diketahui masing-masing  $P(V_j)$  dan  $P(W_k|V_j)$  kemudian dihitung dengan perhitungan metode *Naïve Bayes* untuk setiap klasifikasinya. Perhitungan ini mengambil kata-kata dari basis data yang sudah dibuat sebelumnya. Jika dalam data contoh terdapat kata yang sama dengan di basis data, maka kata tersebut diambil untuk kemudian dimasukkan persamaan. Metode *Naïve Bayes* seperti berikut:

**Tabel 2.1. Hasil Proses Pembelajaran (1)**

Kategori	$P(V_j)$	$P(W_k V_j)$		
		jujur	ceria	ramah
Sanguinis	$\frac{1}{4}$	$\frac{1}{200}$	$\frac{2}{200}$	$\frac{2}{200}$
Koleris	$\frac{1}{4}$	$\frac{1}{200}$	$\frac{1}{200}$	$\frac{1}{200}$
Melankolis	$\frac{1}{4}$	$\frac{1}{200}$	$\frac{1}{200}$	$\frac{1}{200}$
Plegmatis	$\frac{1}{4}$	$\frac{1}{200}$	$\frac{1}{200}$	$\frac{2}{200}$

**Tabel 2.2. Hasil Proses Pembelajaran (2)**

Kategori	P(V <sub>j</sub> )	P(W <sub>k</sub>  V <sub>j</sub> )		
		sabar	humoris	gaul
Sanguinis	$\frac{1}{4}$	$\frac{1}{200}$	$\frac{2}{200}$	$\frac{1}{200}$
Koleris	$\frac{1}{4}$	$\frac{1}{200}$	$\frac{1}{200}$	$\frac{1}{200}$
Melankolis	$\frac{1}{4}$	$\frac{1}{200}$	$\frac{1}{200}$	$\frac{1}{200}$
Plegmatis	$\frac{1}{4}$	$\frac{2}{200}$	$\frac{1}{200}$	$\frac{2}{200}$

**Tabel 2.3. Hasil Proses Pembelajaran (3)**

Kategori	P(V <sub>j</sub> )	P(W <sub>k</sub>  V <sub>j</sub> )		
		tualang	kenal	sedih
Sanguinis	$\frac{1}{4}$	$\frac{1}{200}$	$\frac{1}{200}$	$\frac{1}{200}$
Koleris	$\frac{1}{4}$	$\frac{2}{200}$	$\frac{1}{200}$	$\frac{1}{200}$
Melankolis	$\frac{1}{4}$	$\frac{1}{200}$	$\frac{1}{200}$	$\frac{1}{200}$
Plegmatis	$\frac{1}{4}$	$\frac{1}{200}$	$\frac{1}{200}$	$\frac{1}{200}$

Kemudian hasil perkalian yang diperoleh dari persamaan 1 dan 2 dilihat nilai perhitungan terbesar di antara keempat kelas. Nilai terbesar merupakan kesimpulan dari klasifikasi teks. Dalam hal ini ada dua nilai yang sama besar yaitu sanguinis dan plegmatis. Maka dokumen contoh yang dimasukkan diawal tersebut termasuk dokumen data kepribadian termasuk dalam klasifikasi sanguinis dan plegmatis (Lestari, Putra, & Chayawan, 2013). Seperti perhitungan berikut:

$$\begin{aligned}
 P(\text{sanguinis}|\text{dokumen}) &= P(V_j) \times P(W_k|V_j) \\
 &= \frac{1}{4} \times \frac{1}{200} \times \frac{2}{200} \times \frac{2}{200} \times \frac{1}{200} \times \frac{1}{200} \\
 &\quad \times \frac{2}{200} \times \frac{1}{200} \times \frac{1}{200} \times \frac{1}{200} \times \frac{1}{200} \\
 &= \underline{3,09 \times 10^{-21}}
 \end{aligned}$$

$$\begin{aligned}
 P(\text{koleris}|\text{dokumen}) &= P(V_j) \times P(W_k|V_j) \\
 &= \frac{1}{4} \times \frac{1}{200} \times \frac{1}{200} \times \frac{1}{200} \times \frac{1}{200} \times \frac{1}{200} \\
 &\quad \times \frac{1}{200} \times \frac{1}{200} \times \frac{2}{200} \times \frac{1}{200} \times \frac{1}{200} \\
 &= 0,977 \times 10^{-21}
 \end{aligned}$$

$$\begin{aligned}
 P(\text{melankolis}|\text{dokumen}) &= P(V_j) \times P(W_k|V_j) \\
 &= \frac{1}{4} \times \frac{1}{200} \times \frac{1}{200} \times \frac{1}{200} \times \frac{1}{200} \times \frac{1}{200} \\
 &\quad \times \frac{1}{200} \times \frac{1}{200} \times \frac{1}{200} \times \frac{1}{200} \times \frac{1}{200} \\
 &= 0,488 \times 10^{-21}
 \end{aligned}$$

$$\begin{aligned}
 P(\text{plegmatis}|\text{dokumen}) &= P(V_j) \times P(W_k|V_j) \\
 &= \frac{1}{4} \times \frac{1}{200} \times \frac{2}{200} \times \frac{2}{200} \times \frac{1}{200} \times \frac{1}{200} \\
 &\quad \times \frac{2}{200} \times \frac{1}{200} \times \frac{1}{200} \times \frac{1}{200} \times \frac{1}{200} \\
 &= \underline{3,09 \times 10^{-21}}
 \end{aligned}$$

### 2.3.2. Evaluasi Metoda

Dalam klasifikasi teks juga perlu mengetahui apakah penggunaan metode sudah tepat atau belum. Ada beberapa

pengujian untuk mengevaluasi keefektifan metode dalam menyelesaikan persoalannya. Untuk mengukurnya menggunakan uji akurasi, uji presisi, dan uji *recall*. Ikonomakis, Kotsiantis, & Tampakas (2005) menyatakan bahwa untuk menghitung pengujian tersebut akan lebih mudah bila memahami apakah data termasuk *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), atau *False Negative* (FN), *Predicted Positive* (PP), atau *Real Positive* (RP) seperti ditunjukkan tabel 2.4.

**Tabel 2.4. Klasifikasi Dokumen**

Singkatan	Definisi
TP	Ditentukan sebagai sebuah dokumen yang diklasifikasikan dengan benar dan sesuai dengan kategorinya
FP	Ditentukan sebagai sebuah dokumen yang diklasifikasikan dengan salah dan sesuai dengan kategorinya
FN	Ditentukan sebagai sebuah dokumen yang dianggap salah oleh mesin namun kenyataannya tidak sesuai dengan kategori
TN	Ditentukan sebagai sebuah dokumen yang salah dan kenyataannya juga tidak sesuai dengan kategori
PP	Data yang terhitung kategori positif oleh mesin pembelajaran meskipun dalam kenyataan adalah negatif. Ini digunakan pada perhitungan uji presisi
RP	Data yang dalam kenyataan merupakan hasil positif meskipun oleh mesin pembelajaran dinyatakan negatif. Ini digunakan pada perhitungan uji <i>recall</i>

Pengujian data dari mesin klasifikasi teks dilakukan dengan uji akurasi, uji presisi, dan *recall test* dengan dibandingkan pada hasil klasifikasi secara manual seperti ditunjukkan tabel 2.5. Pada dasarnya, baik uji akurasi, uji presisi, dan uji *recall*, nilainya berada pada rentang 0 s/d 1. Oleh karena itu, suatu program klasifikasi teks yang baik adalah yang dapat memberikan nilai yang mendekati 1. Nilai presisi dan nilai uji *recall* dianggap sebagai pengujian untuk mengetahui keberhasilan sistem, sedangkan uji akurasi adalah untuk mengetahui keberhasilan proses klasifikasi secara umum. Sub bab berikutnya menjelaskan mengenai perhitungan uji akurasi, uji presisi, dan uji *recall*.

**Tabel 2.5. Tabel Kontingensi**

		Manual (kenyataan)		
		Positive	Negative	
Mesin (prediksi)	Positive	TP	FP	PP
	Negative	FN	TN	
		RP		

Uji akurasi ( $A_i$ ) adalah ukuran yang paling sering digunakan untuk teknik kategorisasi. Nilai akurasi memperhatikan keseluruhan data dokumen sehingga lebih bervariasi jumlah yang benar daripada presisi dan *recall* seperti pada persamaan 3.

$$A_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \quad (3)$$

Uji presisi ( $\Pi_i$ ) adalah menentukan kondisi kemungkinan sebuah dokumen acak diklasifikasikan pada kategori atau termasuk dalam kategori yang benar. Ini mengartikan dokumen adalah benar oleh sebuah mesin pembelajaran sehingga pengujiannya memperhatikan kebenaran dalam kenyataannya. Persamaan uji presisi ditunjukkan pada persamaan 4.

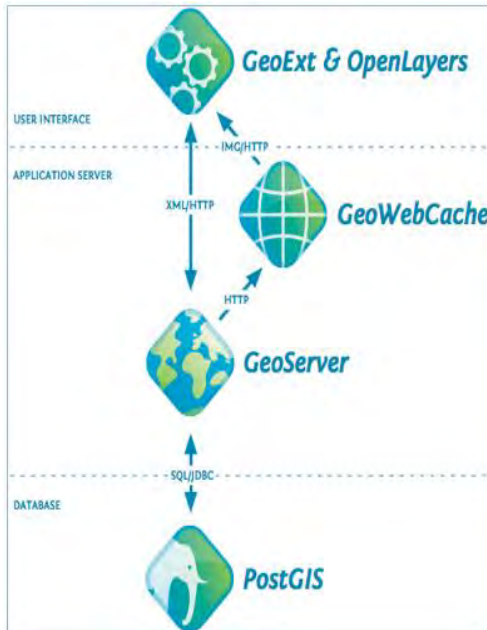
$$\pi_i = \frac{TP_i}{TP_i + FP_i} \quad (4)$$

*Recall test* ( $\rho_i$ ) adalah mendefinisikan sebagai kemungkinan jika sebuah dokumen acak harus diklasifikasikan dalam kategori. Ini mengartikan memperhatikan dokumen yang pada dasarnya sesuai dengan kategori baik itu benar atau salah dalam klasifikasi oleh mesin pembelajaran sebagaimana ditunjukkan pada persamaan 5.

$$\rho_i = \frac{TP_i}{TP_i + FN_i} \quad (5)$$

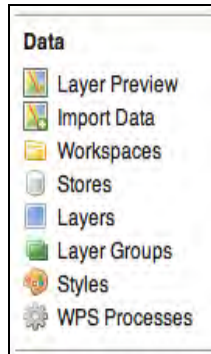
## 2.4. Aplikasi OpenGeo Suite

Aplikasi OpenGeo Suite adalah seperangkat aplikasi mengenai pemetaan data yang lisensinya diberikan kepada pengguna secara bebas. Aplikasi ini merupakan sebuah platform *geospatial* yang terdiri beberapa aplikasi antara lain: PostGIS, GeoServer, GeoWebCache, OpenLayers, QGIS, dan GeoExt (lihat gambar 2.4). Keseluruhan aplikasi tersebut telah terintegrasi, sehingga memudahkan pengguna dalam mengolah data mentah, hingga dipetakan kedalam *heat map*. OpenGeo Suite adalah aplikasi *open source* yang dibuat oleh perusahaan yang berdomisili di Amerika Serikat, BoundlessGeo.



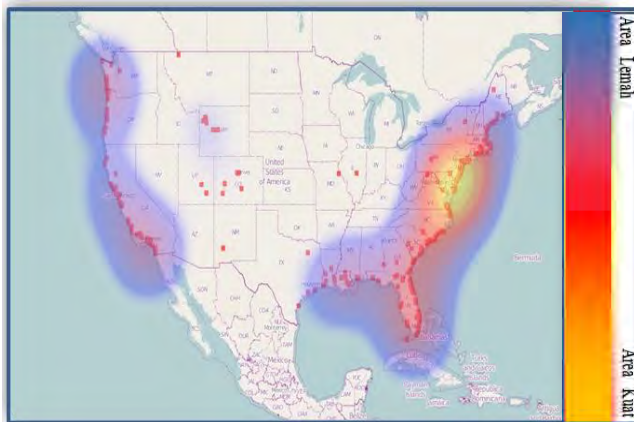
**Gambar 2.4. Arsitektur OpenGeo Suite  
(BoundlessGeo, 2014)**

Cara kerja OpenGeo Suite awalnya memasukkan data kedalam *data storage*, yang mana dalam hal ini menggunakan aplikasi PostGIS. Kemudian data tersebut diolah di lapisan *application server*, menggunakan aplikasi GeoServer. Sedangkan penentuan *layer* untuk menampilkan hasil pengolahan data, menggunakan aplikasi GeoWebCache. Akhirnya aplikasi dapat ditampilkan oleh aplikasi GeoExt & OpenLayer. Aplikasi-aplikasi ini sudah saling terintegrasi, sehingga data yang dimasukkan kedalam *data storage* dapat langsung diolah melalui berbagai proses hingga menjadi aplikasi baru. Fitur-fitur OpenGeo Suite ditunjukkan pada gambar 2.5.



**Gambar 2.5. Fitur-fitur OpenGeo Suite  
(BoundlessGeo, 2014)**

Hasil dari pengolahan data, berdasarkan arsitektur OpenGeo Suite, nantinya akan menampilkan pemetaan data dalam bentuk *heat map* (lihat gambar 2.6). *Heat map* ini menunjukkan *clustering* yang memiliki kemiripan data. Semakin banyak data yang memiliki suatu kemiripan diberikan warna yang berbeda sesuai pengaturan di fitur *styles*.



**Gambar 2.6. Tampilan *Heat Map***



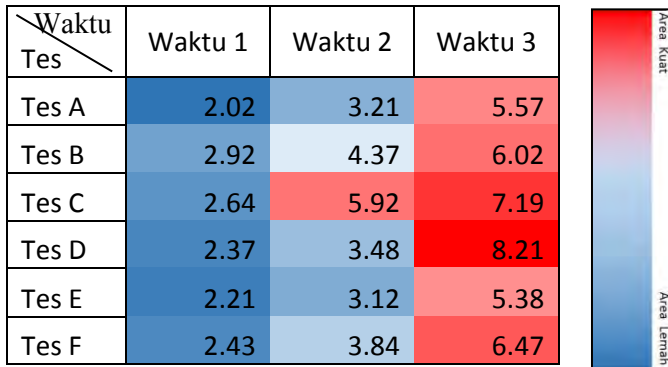
Semakin banyak data dalam suatu area, maka grafis pada peta akan menampilkan warna biru hingga kemerahan. Sedangkan jika hanya ada satu data dalam suatu area, maka hanya akan ditampilkan dalam bentuk node merah dengan sedikit warna grafis biru di sekitarnya (BoundlessGeo, 2014).

## 2.5. *Heat Map*

*Heat map* adalah cara grafis yang menampilkan data-data daritabel dengan menggunakan warna untuk merepresentasikan nilai numerik. Algoritma *clustering* relasi antara baris atau kolom yang memiliki kemiripan. Sebagai contoh, data dengan nilai rendah cenderung lebih ke warna yang lebih dingin, atau secara alamiah diberi warna biru. Sedangkan nilai tinggi lebih ke warna yang lebih panas, atau lebih cenderung ke warna jingga atau merah. Sehingga, semakin merah warna grafis, artinya mengindikasikan bahwa kepadatan data, atau data yang memiliki kemiripan, semakin tinggi. Sebaliknya, jika semakin sedikit kemiripan antar data, maka akan diinterpretasikan dengan warna kuning kebiruan (Vienna, 2013)

*Heat map* dapat dibuat dengan menggunakan teknik *hierarchical clustering*, yang mana dilakukan dengan cara menyusun data dalam hirarki berdasarkan jarak atau kemiripan antar data. Hasil dari perhitungan *hierarchical clustering* ditampilkan dalam bentuk *heat map* sebagai sebuah dendrogram. Umumnya, untuk hal ini, tipe *cluster similarity* yang digunakan adalah *max linkage* atau *complete linkage* (Vienna, 2013). Pendekatan ini berdasarkan dua kemiripan poin terjauh dalam *cluster* yang berbeda. Cara ini dipilih karena memiliki kelebihan dalam mengurangi kemungkinan terjadinya *noise* dan *outlier* (Tan, Steinbach, & Kumar, 2006). Namun konsep *single linkage* atau *average linkage* juga bisa digunakan untuk *heat map*. Sehingga *heat map* dapat menggunakan *single linkage*, *average linkage*, atau *complete linkage* sesuai dengan kebutuhan pembuatan klasterisasi.

Gambar 2.7 menunjukkan intepretasi warna *heat map*. Dalam contoh sederhana tersebut, semakin rendah nilai tes maka semakin berwarna biru, sedangkan jika semakin tinggi nilai tes maka semakin berwarna merah. Data tersebut terdapat atribut waktu 1, waktu 2, dan waktu 3. Atribut tersebut dapat dijadikan sebagai koordinat x, y, dan z. Semakin tinggi nilai suatu atribut, maka warnanya semakin merah. Lalu dihitung jarak tiap-tiap data contoh (*instance*), dapat menggunakan *Manhattan Distance*, yakni perhitungan dengan memperhatikan selisih antar koordinat. Kemudian dilakukan perhitungan *cluster similarity* berdasarkan *complete linkage*.

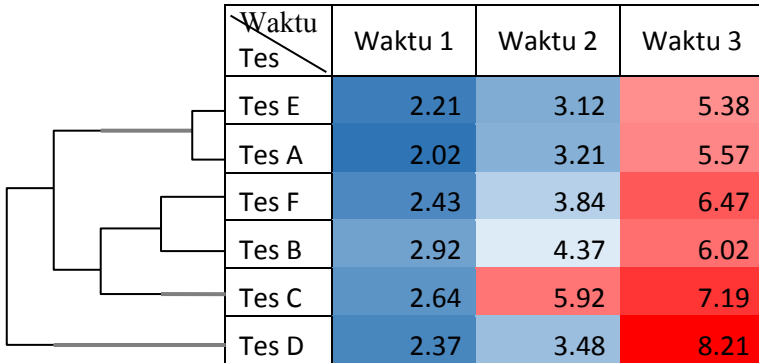


**Gambar 2.7. Contoh Intepretasi Warna**

Setelah melalui perhitungan *cluster similarity*, maka dapat digambarkan dendrogram hasil dari perhitungan tersebut seperti ditunjukkan gambar 2.8. Hasil dari *hierarchical clustering* ini dapat diintepretasikan ke dalam warna menjadi sebuah *heat map* (Thomson Reuters Life Sciences, 2014). Untuk gradasi warna berdasarkan besarnya nilai dapat dilihat pada gambar 2.8.

Cara seperti ini juga diterapkan oleh aplikasi OpenGeo Suite dalam membuat *heat map*. Namun antara hasil contoh tersebut dengan hasil penggunaan aplikasi OpenGeo Suite yang

sedikit membedakan adalah adanya atribut koordinat yang akan dipetakan yaitu garis bujur (*longitude*) dan garis lintang (*latitude*).



**Gambar 2.8.** Dendrogram *Hierarchical Clustering*

## 2.6. *Customer Relationship Management (CRM)*

*Customer Relationship Management (CRM)* adalah kombinasi dari manusia, proses, dan teknologi untuk memahami pelanggan. CRM adalah sebuah pendekatan terintegrasi untuk mengendalikan hubungan dengan berfokus pada menjaga pelanggan dan pengembangan hubungan. CRM dapat membantu organisasi memajemen interaksi pelanggan lebih efektif untuk menjaga kompetisi dalam persaingan ekonomi. Saat ini organisasi-organisasi mengadopsi CRM sebagai strategi bisnis perusahaan. Berdasarkan aspek bagaimana pelanggan berinteraksi dengan organisasi, menurut Mishra (2009) CRM dibedakan menjadi tiga area sbb:

- Operasional CRM mendukung proses *front office* perusahaan seperti *call center*. Keluhan-keluhan pelanggan yang diterima oleh *call center* termasuk dalam area operasional CRM. Opini pelanggan yang

disampaikan harus berdasarkan layanan yang diterima secara langsung yang erat kaitannya dengan hal teknis.

- Analitis CRM membangun informasi apa yang ada dari operasional CRM seperti segmen pelanggan, tingkah laku, dan nilai pelanggan menggunakan metode statistik. Area ini muncul setelah ada analitis dari pelanggan terhadap suatu kebijakan organisasi.
- Kolaboratif CRM berkonsentrasi pada integrasi pelanggan melalui beberapa hubungan interaksi, contohnya toko online. Banyak organisasi menggunakan portal dalam komunikasinya dengan pelanggan.

## 2.7. Operasional CRM PT. Telkomsel

PT. Telkomsel adalah penyedia komunikasi telepon genggam dan layanan data digital terbesar di Indonesia. Telkomsel telah memimpin industri seluler di Indonesia yang saat ini bertransformasi ke era digital. Berdasarkan *Annual Report* 2012, Telkomsel menciptakan komitmen dengan membuat program Telkomsel 2.0 (gambar 2.10). Ini adalah bentuk upaya pengelolaan hubungan pelanggan dengan transformasi bisnis dan sumber daya manusia untuk mencapai tujuan Telkomsel dan dalam menghadapi tantangan bisnis. Telkomsel bertujuan untuk memimpin pasar dalam pelayanan digital yaitu memberikan solusi memenuhi ekspektasi pelanggan dan menciptakan nilai ke semua pemangku saham.



Gambar 2.9. Logo PT. Telkomsel

Inti dari transformasi Telkomsel 2.0 adalah ekosistem *Device-Network-Application* (D-N-A) yang secara konstan ditingkatkan untuk menyediakan pengalaman telekomunikasi *end-to-end* terbaik di Indonesia. Sehingga ini memerlukan penanganan atau pengelolaan *feedback* dari pelanggan untuk dianalisis agar dapat menciptakan hubungan dengan pelanggan yang baik.

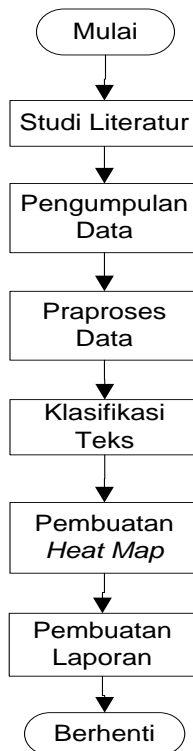


**Gambar 2.10. Logo Telkomsel 2.0**

### BAB III

## METODOLOGI Pengerjaan Tugas Akhir

Metodologi pengerjaan tugas akhir diawali dari studi literatur, lalu dilanjutkan dengan pengambilan data Twitter. Data yang terkumpul dilakukan praproses untuk memastikan agar data dapat diproses selanjutnya. Kemudian data hasil seleksi dilakukan klasifikasi teks hingga kemudian dipetakan ke dalam *heat map*. Secara umum diagram alur metodologi pengerjaan tugas akhir ditunjukkan pada gambar 3.1.



Gambar 3.1. Metodologi Pengerjaan Tugas Akhir

### 3.1. Studi Literatur

Studi literatur yang dilakukan adalah pembelajaran dan pemahaman literatur yang berkaitan dengan permasalahan yang ada. Beberapa hal yang dipelajari adalah Twitter API *Stream* untuk mengumpulkan data, pembuatan klasifikasi teks data Twitter dengan menggunakan metode Naive Bayes, dan mengoperasikan OpenGeo Suite agar data Twitter yang sudah dilakukan klasifikasi teks dapat menghasilkan peta dalam bentuk *heat map*.

Untuk pembejalaran mesin tentang klasifikasi teks, literatur yang digunakan adalah jurnal milik Lestari, Putra, dan Chayawan yang dipublikasikan di *International Journal of Computer Sciences Issues* (IJCSI). Di dalamnya dijelaskan teknis melakukan klasifikasi teks menggunakan metode Naive Bayes. Sedangkan untuk pemahaman mengenai klasifikasi teks itu sendiri, merujuk pada penjelasan jurnal milik Bing Liu. Kemudian untuk mengoperasikan aplikasi OpenGeo Suite, referensi yang digunakan bersumber dari situs BoundlessGeo di halaman panduan. Situs ini juga menyediakan aplikasi OpenGeo Suite secara *open source*. Sedangkan untuk teknis melakukan *heat map* dari aplikasi tersebut, rujukan yang digunakan bersumber dari website Vienna, dan Thomson Reuters Life Science, juga literatur milik Steinbach, Tan, dan Kumar.

### 3.2. Pengumpulan Data

Tahap ini dilakukan pengumpulan data melalui Twitter API *Stream* dengan bahasa pemrograman *python*. Proses dilakukan dengan memanfaatkan program berupa *open source* milik perusahaan SentDex. SentDex juga menyediakan panduan mengenai cara penggunaan program tersebut, [sentdex.com](http://sentdex.com). Namun dalam panduan tersebut ada beberapa langkah yang dilewati seperti tidak dijelaskan *packages* apa saja yang perlu di install, dsb. Untuk permasalahan ini *package* yang perlu di install adalah *Oauth2* (lihat Bab 2 Tinjauan Pustaka Sub bab 2.1 Twitter API *Stream*). Ini digunakan sebagai autentikasi program Twitter API *Stream* untuk dapat mengambil data Twitter.

### 3.2.1. Persiapan Twitter API *Stream*

Twitter API *Stream* adalah teknik pengambilan data Twitter yang prosesnya dilakukan secara waktu nyata. Untuk menjalankan Twitter API *Stream* memerlukan *token* API. Cara memperolehnya adalah dengan mendaftar sebagai pengembang aplikasi Twitter di [dev.twitter.com](https://dev.twitter.com). Kemudian setelah mengisi formulir, *token-token* yang diperlukan akan muncul seperti *consumer key*, *consumer secret*, *access token*, dan *access token secret* seperti ditunjukkan gambar 3.2. Keempat nomor unik inilah yang digunakan Twitter API *Stream* sebagai hak ijin untuk menggali data Twitter. Paket Twitter API *Stream* berfungsi untuk menjalankan proses pengambilan data Twitter, sedangkan paket otorisasi adalah *token* untuk mengenalkan Twitter agar diperkenankan untuk mengambil datanya.



The image shows a screenshot of the 'OAuth Settings' page. It contains four input fields, each with a unique alphanumeric string and a redacted portion. The fields are: 'Consumer key: \*' with value 'FmnCyfAWiDagNUXt3...', 'Consumer secret: \*' with value '0XxDqeOIPt874mJhja8CDMD5YOY28ABwMipnD...', 'Access token:' with value '116125397-dDygcdlB0p4V898UOj7phT6wFSg43nCKY...', and 'Access token secret:' with value 'bYZVRP4NiiMoOwaNI5jt8KkAe3CFW051zATiKXC...'. Below the second and fourth fields, there is a note: 'Remember this should not be shared.'

Gambar 3.2. Nomor *Token* Twitter API



Kemudian memasukkan nomor-nomor *token* ke dalam program Twitter API *Stream*. Nomor *token* ini hanya dapat diperoleh dengan mendaftar sebagai pengembang Twitter. Jika ada dua program Twitter API *Stream* menggunakan nomor *token* yang sama, maka hanya salah satu program yang lebih dulu mengakses yang dapat menggunakan nomor *token* tersebut. Program kedua tidak dapat masuk karena nomornya sudah digunakan. Oleh karena itu, kerahasiaan nomor unik ini perlu diperhatikan. Pada pengerjaan tugas akhir ini, penggunaan program Twitter API *Stream* tidak menggunakan *proxy* melainkan akan terhubung dengan koneksi modem. Jika ingin mengakses melalui *proxy*, maka ada pengaturan *proxy* terlebih dahulu

Ada dua *packages* yang memiliki peran utama dalam menjalankan program Twitter API *Stream*. Kedua *packages* tersebut adalah *python-oauth2-master* dan *tweepy-master*. Keduanya memiliki fungsi yang berbeda-beda. *Python-oauth2-master* berguna untuk mengaktifkan otorisasi nomor *token* Twitter API agar dapat dikenali oleh Twitter. Sedangkan *tweepy-master* berguna agar program Twitter API *Stream* dapat terus berjalan secara berulang dalam menggali data sesuai dengan kata kunci yang dimasukkan.

Apabila paket *python-oauth2-master* dan paket *tweepy-master* telah diinstall dan nomor *token* telah dimasukkan, maka proses pengumpulan data Twitter dapat dijalankan. Proses penggalian data Twitter akan terus berjalan hingga diberhentikan secara manual. Proses ini berlangsung secara waktu nyata artinya data Twitter yang diperoleh berdasarkan data yang dilayangkan pada saat itu sehingga bukan data historis. Proses ini terus berjalan selama masih ada akun yang melayangkan *tweet* dengan muatan kata ‘telkomsel’. Jika dirasa data sudah cukup, proses dapat diberhentikan dengan cara menyimpan file tersebut (Ctrl+S), maka file akan tersimpan dalam ekstensi json.

### 3.2.2. Proses Twitter API Stream

*Token* autentikasi disalin ke program Twitter API Stream. Data yang diambil adalah *tweet* yang mengandung kata 'telkomsel'. Dengan program ini, *tweet* yang akan diambil adalah *tweet* yang mengandung *mention* (@telkomsel), *hashtag* (#telkomsel), maupun teks. Proses pengambilan data dilakukan selama beberapa hari hingga jumlah data dirasa cukup representatif untuk dipetakan. Kemudian jika sudah, program Twitter API Stream dapat dihentikan, lalu hasilnya disimpan dalam bentuk file. Lihat gambar 3.3 untuk ilustrasi proses Twitter API Stream.

Proses pengumpulan data Twitter berlangsung berdasarkan kata kunci yang telah ditetapkan sebelum program Twitter API Stream dijalankan. Maka data *tweet* yang diambil hanyalah yang mengandung kata kunci tersebut. Proses ini mengambil data *tweet* secara waktu nyata. Data yang sudah lampau atau data historis tidak dapat diambil kembali. Proses pengumpulan data Twitter baru akan berhenti ketika proses Twitter API Stream dihentikan.

Pada langkah-langkah berikutnya, data dengan ekstensi json tersebut sulit untuk diolah. Oleh karena itu, data perlu disesuaikan program klasifikasi teks yang dapat diolah file ekstensi tertentu seperti csv, html, dsb. Untuk dapat melakukan itu, data perlu dikonversi dari file ekstensi json ke csv atau html. Setelah itu data siap untuk dilakukan praproses data.

Sebuah data *tweet* yang diambil memiliki banyak atribut dengan nilai yang berbeda-beda. Atribut yang tidak bersesuaian atau tidak diperlukan perlu dihapuskan. Oleh karena itu, proses selanjutnya dilakukan praproses data untuk memastikan bahwa data yang akan diklasifikasikan dan dipetakan memiliki atribut yang dibutuhkan saja.

```

Python Shell*
File Edit Shell Debug Options Windows Help
{"created_at":"Thu May 29 01:25:08 +0000 2014","id":471824527354109952,"id_str":"471824527354109952","text":"Lelet bener jaringan Telkomsel disini hft .","source":"\u003ca href=\"http://blackberry.com/twitter\" rel=\"nofollow\" \u003eTwitter for BlackBerry\u003c\/a\u003e","truncated":false,"in_reply_to_status_id":null,"in_reply_to_status_id_str":null,"in_reply_to_user_id":null,"in_reply_to_user_id_str":null,"in_reply_to_screen_name":null,"user":{"id":636919284,"id_str":"636919284","name":"Aulia\u265b","screen_name":"Auliarhmdhni","location":"","url":"http://ask.fm/Auliaramadhan123","description":"I want one direction concert tickets!!! Pest a dan b","protected":false,"followers_count":728,"friends_count":768,"listed_count":1,"created_at":"Mon Jul 16 09:36:50 +0000 2012","favourites_count":241,"utc_offset":null,"time_zone":null,"geo_enabled":true,"verified":false,"statuses_count":9461,"lang":"id","contributors_enabled":false,"is_translator":false,"is_translation_enabled":false,"profile_background_color":"252B33","profile_background_image_url":"http://pbs.twimg.com/profile_background_images/826738407/a8141c4852e24db03c7039e376228131.jpeg","profile_background_image_url_https":"https://pbs.twimg.com/profile_background_images/826738407/a8141c4852e24db03c7039e376228131.jpeg","profile_background_tile":true,"profile_image_url":"http://pbs.twimg.com/profile_images/458260135001878528/68ESsqq_normal.jpeg","profile_image_url_https":"https://pbs.twimg.com/profile_images/458260135001878528/68ESsqq_normal.jpeg","profile_banner_url":"https://pbs.twimg.com/profile_banners/636919284/1401290745","profile_link_color":"D02B55","profile_sidebar_border_color":"FFFFFF","profile_sidebar_fill_color":"DDEEF6","profile_text_color":"333333","profile_use_background_image":false,"default_profile":false,"default_profile_image":false,"following":null,"follow_request_sent":null,"notifications":null},"geo":{"type":"Point","coordinates":[-6.16467,106.83616]},{"type":"Point","coordinates":[106.83616,-6.16467]},{"id":"25f5e9dbc1b635ac","url":"https://api.twitter.com/1.1/geo/id/25f5e9dbc1b635ac.json","place_type":"city","name":"Gambir","full_name":"Gambir, Jakarta Pusat","country_code":"ID","country":"Indonesia","contained_within":[],"bounding_box":{"type":"Polygon","coordinates": [[[106.807441,-6.180635],[106.807441,-6.157287],[106.846151,-6.157287],[106.846151,-6.180635]]]},"attributes":{"contributors":null,"retweet_count":0,"favorite_count":0,"entities":{"hashtags":[],"symbols":[],"urls":[],"user_mentions":[]},"favorited":false,"retweeted":false,"filter_level":"medium","lang":"in"}}

{"created_at":"Thu May 29 01:25:30 +0000 2014","id":471824622531674113,"id_str":"471824622531674113","text":"@Telkomsel saat ini lagi nyantai di luar ruangan nih. Signal GSM nya full koq, cuma gak ada 3G atau H nya. Kemarin m alam masih bisa.","source":"\u003ca href=\"https://mobile.twitter.com\" r
Ln: 11 | Col: 0

```

Gambar 3.3. Ilustrasi Proses Pengumpulan Data Twitter

### 3.3. Praproses Data

Data mentah yang diperoleh masih memerlukan pemrosesan untuk mengubahnya menjadi data yang siap digunakan. Proses-proses yang dilakukan dalam tahap praproses data adalah pemilihan atribut data, eliminasi data tanpa koordinat, pemilahan data opini dan non-opini, serta pemilihan opini yang merupakan operasional CRM. Untuk tahap praproses data ini, semua langkah-langkahnya dilakukan secara manual.

#### 3.3.1. Pemilihan Atribut Data

Data mentah awalnya dipilih atribut-atribut yang bersesuaian kebutuhan dalam pemetaan. Untuk itu data mentah harus mempunyai koordinat bujur dan lintang. Selain itu data tersebut juga mengalami klasifikasi sentimen sehingga atribut teks juga diambil. Data yang baru diperoleh tersebut, selain koordinat, juga memiliki banyak atribut lain seperti *in\_reply\_in\_status*, *lang*, *verified*, *is\_translator*, dll. Menurut Ghiassi, Skinner, & Zimbra (2013) total ada sebanyak 195 atribut. Atribut-atribut inilah yang kurang informatif untuk ditampilkan ke peta nantinya. Untuk itulah atribut-atribut itu dihapus, dan dipilih atribut-atribut lain yang bersesuaian dan lebih informatif, seperti atribut *text*, *name*, dll. Proses selanjutnya adalah menghilangkan tanda baca yang tidak relevan. Data Twitter yang diperoleh memiliki ciri-ciri panjang maksimal 140 karakter, singkatan, akronim, *emoticon*, *hashtag* (#), referensi pengguna, dan link yang tertempel (Ghiassi, Skinner, & Zimbra, 2013).

#### 3.3.2. Eliminasi Data tanpa Koordinat

Dalam melakukan pemetaan, data membutuhkan atribut yang memiliki koordinat bujur dan lintang. Namun tidak semua *tweet* mengandung koordinat. Ini dikarenakan saat pengaturan akun memang tidak mengaktifkan koordinatnya. Oleh karena itu, data yang terkumpul dipilih yang memiliki koordinat. Jika tidak ada koordinat maka data harus dieliminasi.

### 3.3.3. Pemilahan Data Opini dan Non-Opini

Data yang memiliki koordinat dapat berupa opini atau non-opini. Oleh karena itu, data tersebut kemudian dipilah-pilah untuk memastikan data yang diklasifikasikan adalah data opini. Karena konsep klasifikasi teks adalah memaknai data teks ke dalam suatu kelas. Sehingga data non-opini harus dieliminasi dari pemrosesan selanjutnya. Tahapan pemilahan data opini dan non-opini merupakan tahap yang hasilnya dapat mempengaruhi akurasi metode *Naïve Bayes*. Ini karena pernyataan yang terkandung di dalam *tweet* non-opini berupa fakta, berita, iklan, atau pendapat yang tidak bermakna sehingga tidak dapat mewakili keinginan pelanggan. Data yang terambil selama proses pengumpulan data cukup banyak mengandung data berjenis non-opini. Sedangkan data opini dibedakan menjadi dua jenis yaitu opini positif dan negatif. Contoh masing-masing pernyataan ditunjukkan pada tabel 3.1.

Tabel 3.1. Pemilahan Data Opini

Jenis	Contoh Tweet
Non-opini	<i>“Layanan Favorite Digital - Products - Telkomsel “</i>
	<i>“@Telkomsel udah. Trus?”</i>
	<i>“Bagi anda warga Kota Bogor bisa gunakan Layanan SMS Kepolisian, Format: Bogor &lt;spasi&gt; Isi SMS. Kirim ke 9123 (Sementara hanya Telkomsel)”</i>
Opini negatif	<i>“@Telkomsel : Kecewa. Tarif naik tp malah makin ga bagus koneksinya. Signal ga stabil. Mahalnya aja nih.”</i>
Opini positif	<i>“Love love lah jaringan telkomsel kalo macam nii hahaha”</i>

### 3.3.4. Pemilihan Data Opini dengan Aspek Operasional CRM

Berdasarkan pada sub bab 1.5. relevansi atau manfaat tugas akhir, manfaat yang diharapkan dari pengerjaan Tugas Akhir ini adalah dapat membantu Telkomsel dalam menerapkan operasional CRM. Untuk itu, opini yang disampaikan pelanggan harus yang berhubungan dengan CRM. Salah satu cabang CRM adalah aspek operasional. Aspek ini meliputi segala masukan pelanggan yang berkaitan dengan faktor teknis. Data yang sudah bersih dan sesuai dengan tujuan tugas akhir sehingga siap untuk diproses selanjutnya.

### 3.4. Klasifikasi Teks

Pada tahap ini dilakukan pendefinisian kata-kata untuk dikategorikan ke dalam sentimen positif atau negatif berdasarkan data yang tersedia. Ada dua fase untuk melakukan hal tersebut, yakni fase pembelajaran dan fase klasifikasi. Fase pembelajaran bertujuan untuk memperoleh data pelatihan berdasarkan data sampel. Sedangkan fase klasifikasi untuk mengolah data baru berdasarkan data pelatihan yang telah diperoleh sebelumnya.

Untuk menerapkan klasifikasi teks, diperlukan program untuk mengimplementasikannya agar lebih memudahkan pada proses selanjutnya. Selain itu juga untuk menghindari kesalahan manusia. Proses metode *Naïve Bayes* tersebut dapat dikomputerisasikan menggunakan PHP. Sedangkan untuk menampung kata-kata unik diperlukan basis data yang mana dalam hal ini akan menggunakan MySQL. Program PHP nantinya dapat diintegrasikan dengan MySQL. Sehingga proses pengolahan klasifikasi teks dapat dikomputerisasikan keseluruhan.

Awalnya data sampel yang telah melalui praproses data diolah secara manual untuk dimasukkan kedalam MySQL. Sentimen setiap kata dalam suatu data sampel besar ditentukan sendiri berdasarkan justifikasi hingga kemudian menjadi basis data yang berisikan data pelatihan. Inilah yang dilakukan pada fase pembelajaran. Setelah itu, fase klasifikasi dilakukan saat ada data baru yang ingin diketahui sentimennya. Proses ini memanfaatkan

data pelatihan yang mana telah dihasilkan pada fase pembelajaran. Keluaran yang dihasilkan dari program untuk metode *Naïve Bayes* ini nantinya berupa penghakiman pada tiap-tiap kalimat, yakni positif atau negatif.

Apabila telah melalui proses klasifikasi teks, masing-masing kalimat sudah diketahui sentimennya opini pelanggan terhadap Telkomsel. Sehingga data tersebut telah siap untuk dipetakan ke dalam *heat map* dengan menggunakan aplikasi OpenGeo Suite.

### **3.5. Pembuatan *Heat Map***

Tahap pembuatan *heat map* menggunakan aplikasi OpenGeo Suite, dimana peta yang ditampilkan hanyalah peta Indonesia. Data yang telah melalui proses klasifikasi teks, formatnya diubah menjadi ekstensi teks agar dapat diolah oleh lapisan terbawah OpenGeo Suite yaitu PostGIS. Urutan lapisan PostGIS memungkinkan untuk mengubah koordinat data menjadi sebuah nilai yang mana menjadi suatu titik pada peta. Titik inilah yang nantinya menjadi node. Agar pengolahan berikutnya lebih mudah, data sentimen dibedakan antara klasifikasi positif dan negatif. Kedua data tersebut nantinya akan dibedakan dari warna nodenya.

Suatu data agar dapat digunakan untuk pemetaan harus memiliki atribut koordinat. Kemudian data diproses selanjutnya pada lapisan GeoServer. Pada lapisan ini, data dibuat menjadi *heat map* berdasarkan kepadatan nodenya (lihat Bab 2. Tinjauan Pustaka Sub bab 2.5 *Heat Map*). Ada beberapa tumpukan lapisan untuk membuat suatu kesatuan peta. Tumpukan lapisan tersebut terdiri dari peta Indonesia, pemetaan berdasarkan node, dan pemetaan berdasarkan *heat map*. Setelah itu barulah lapisan-lapisan ditampilkan menjadi satu melalui OpenLayer.

### **3.6. Pembuatan Laporan**

Pada tahapan terakhir ini akan dilakukan pembuatan laporan dalam bentuk buku tugas akhir. Buku ini berisi dokumentasi langkah-langkah pengerjaan tugas akhir secara rinci. Buku ini diharapkan dapat bermanfaat sebagai referensi untuk pengerjaan penelitian lain, serta sebagai acuan untuk pengembangan lebih lanjut terhadap topik pengerjaan tugas akhir yang serupa.



*Halaman ini sengaja dikosongkan*

## **BAB IV DESAIN SISTEM**

Bab ini menjelaskan mengenai desain sistem dari pengklasifikasi teks berbasis *web* dengan menggunakan metode *Naive Bayes*. Bahasa data teks bersumber dari jejaring sosial Twitter, maka kemungkinan ada bahasa tidak baku yang harus diproses oleh aplikasi. Untuk itu aplikasi juga harus didesain agar dapat menyelesaikan persoalan tersebut. Selain itu juga ada desain halaman web yang menampilkan keluaran dari OpenGeo Suite berupa pemetaan data Twitter dalam bentuk *heat map*.

### **4.1. Desain Aplikasi Pengklasifikasi Teks**

Desain aplikasi klasifikasi teks berbasis *web* dengan menggunakan metode *Naive Bayes* meliputi desain basis data, diagram *use case*, diagram sekuen, dan diagram aktivitas

#### **4.1.1. Desain Basis Data Pengklasifikasi Teks**

Basis data klasifikasi teks mempunyai dua tabel yang bernama ‘tweet’ dan ‘keyword’. Ini dikarenakan kedua tabel mempunyai data masukan dengan tujuan yang berbeda. Data masukan untuk tabel ‘tweet’ adalah data dokumen pelatihan, sedangkan data masukan tabel ‘keyword’ adalah data fitur. Data dokumen pelatihan adalah data pelatihan yang berisi kalimat teks Twitter dan koordinatnya, kelas hasil klasifikasi teks, tanggal data teks tersebut dibuat, dan nomor unik (ID). Sedangkan data fitur berisi ID, dan kata-kata kunci dari data dokumen pelatihan beserta nilai skornya. Oleh karena itu, kedua tabel tidak memiliki relasi antara satu dengan lainnya. Data-data tersebut dimasukkan ke basis data MySQL yang dinamakan ‘textmining’. Desain basis data dapat dilihat pada tabel 4.1 dan tabel 4.2

**Tabel 4.1. Desain Basis Data untuk Tabel 'tweet'**

tweet	
PK	<u>id</u>
	date text longitude latitude sentiment

**Tabel 4.2. Desain Basis Data untuk Tabel 'keyword'**

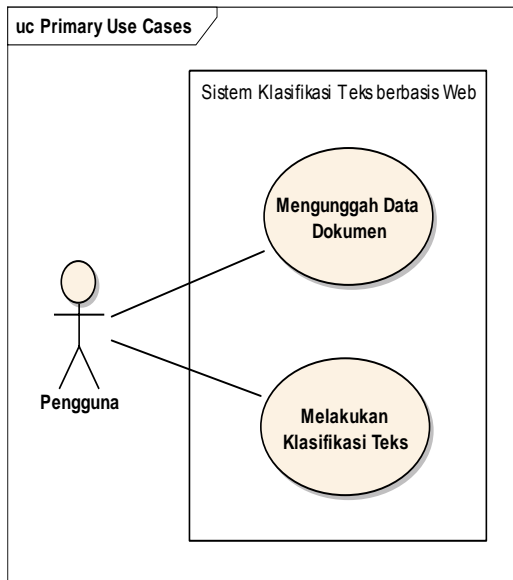
keyword	
PK	<u>idTeks</u>
	teks Pscore Nscore

Tabel 'tweet' menyimpan atribut-atribut yang diperoleh dari pengumpulan data setelah melalui tahapan praproses data. Atribut-atribut ini dibutuhkan oleh OpenGeo Suite untuk melakukan pemetaan. Atribut-atribut tersebut meliputi 'id', 'date', 'text', 'longitude', 'latitude', dan 'sentiment'. Untuk atribut 'sentiment' nilainya kosong karena data masukan belum dilakukan klasifikasi teks. Sedangkan tabel 'keyword' menyimpan kata-kata kunci yang sudah diekstraksi dari data dokumen pelatihan. Atribut 'Pscore' dan 'Nscore' berisi masing-masing frekuensi kata kunci yang muncul dari data dokumen pelatihan.

#### 4.1.2. Diagram *Use Case* Pengklasifikasi Teks

Diagram *use case* berguna untuk menjelaskan fungsi aplikasi yang akan dibangun berdasarkan kebutuhan fungsional. Dalam menggunakan aplikasi klasifikasi teks ini, aktor yang terlibat hanya

pengguna. Fitur aplikasi klasifikasi teks ini terbilang sederhana karena hanya memiliki dua kebutuhan fungsional. Salah satu kebutuhan fungsional melakukan klasifikasi teks bergantung pada kebutuhan fungsional mengunggah data dokumen karena melakukan klasifikasi teks hanya dapat dijalankan jika sudah mengunggah data dokumen. Kedua kebutuhan fungsional dapat dilihat pada gambar 4.1.



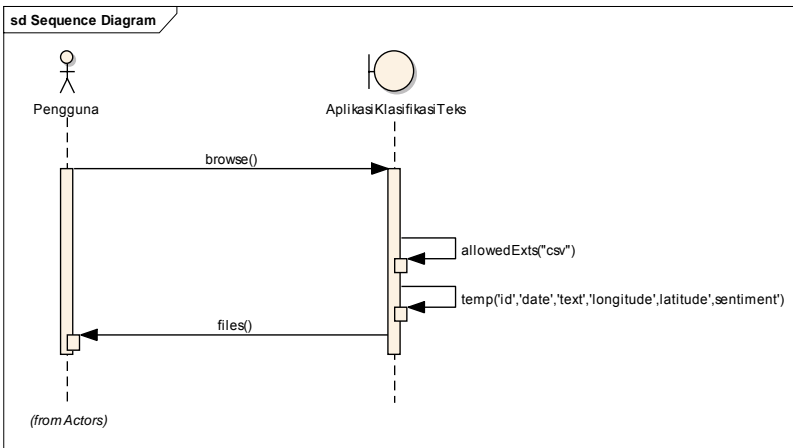
**Gambar 4.1. Diagram *Use Case* Pengklasifikasi Teks**

### **4.1.3. Diagram Sekuen Pengklasifikasi Teks**

Jumlah diagram sekuen sama banyaknya dengan jumlah *use case*. Pada aplikasi klasifikasi teks terdapat dua *use case* sehingga muncul dua diagram sekuen yang dijelaskan pada sub bagian berikutnya.

#### 4.1.3.1. Diagram Sekuen Mengunggah Data Dokumen

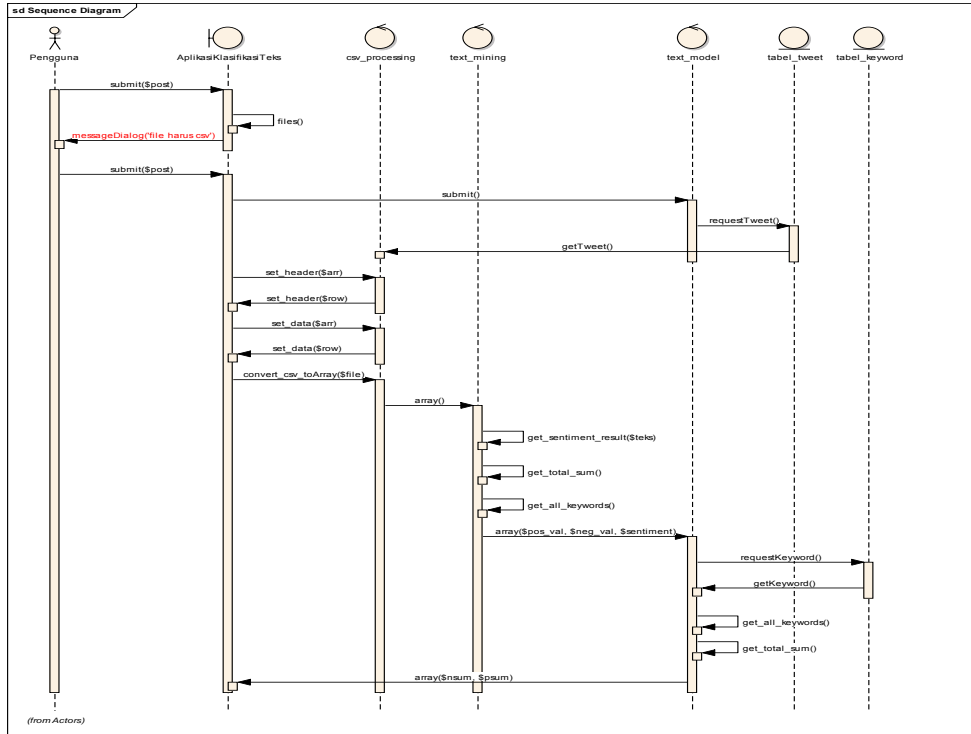
Diagram sekuen ‘mengunggah data dokumen’ merupakan fungsi awal yang dapat dijalankan oleh pengguna. Karena fungsi ‘melakukan klasifikasi teks’ tidak dapat dijalankan jika fungsi tersebut belum dilakukan. Gambar 4.2 menunjukkan diagram sekuen dari *use case* ‘mengunggah data dokumen’. Proses di awali dari mencari file data dokumen di eksplorer. Ada batasan dalam memenuhi fungsi ini yaitu data berformat csv dan *header* file csv berurutan harus ‘id’, ‘date’, ‘text’, ‘longitude’, ‘latitude’, dan ‘sentiment’.



Gambar 4.2. Diagram Sekuen Mengunggah Data Dokumen

#### 4.1.3.2. Diagram Sekuen Melakukan Klasifikasi Teks

Apabila *use case* mengunggah data dokumen telah berhasil dijalankan, maka proses selanjutnya adalah *use case* melakukan klasifikasi teks. Proses ini mengklasifikasikan data yang telah diunggah sebelumnya, yang mana belum diketahui klasifikasinya, lalu dimasukkan ke dalam basis data yang telah disiapkan. Seperti terlihat gambar 4.3



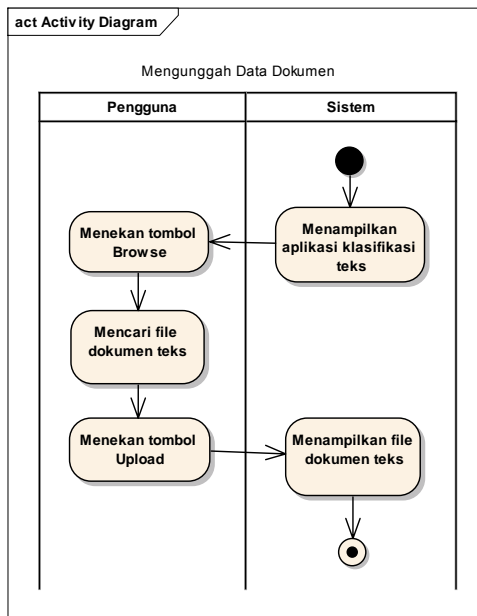
Gambar 4.3. Diagram Sekuen Melakukan Klasifikasi Teks

#### 4.1.4. Diagram Aktivitas Pengklasifikasi Teks

Diagram aktivitas adalah diagram yang menjelaskan mengenai alur proses dari pandangan pengguna dan sistem. Setiap *use case* memiliki diagram aktivitas masing-masing.

##### 4.1.4.1. Diagram Aktivitas Mengunggah Data Dokumen

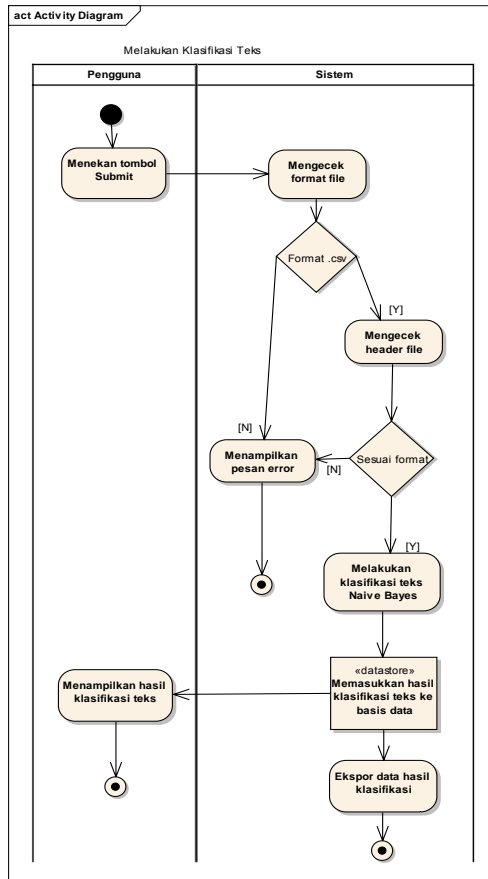
Diagram aktivitas mengunggah data dokumen adalah aktivitas yang harus dilakukan sebelum melakukan klasifikasi teks. Pengguna dapat menekan tombol 'browse' yang sudah ada di aplikasi klasifikasi teks. Seperti terlihat gambar 4.4 kemudian pengguna mencari file dokumen teks yang akan diklasifikasikan. Kemudian menekan tombol 'upload'. Sistem akan menampilkan file dokumen teks yang akan diklasifikasikan.



Gambar 4.4. Diagram Aktivitas Mengunggah Data Dokumen

#### 4.1.4.2. Diagram Aktivitas Melakukan Klasifikasi Teks

Tahapan terakhir pada sistem klasifikasi teks adalah melakukan klasifikasi teks. Pengguna yang sudah mengunggah file dokumen teks dapat menekan tombol 'submit'. Kemudian sistem akan melakukan proses klasifikasi teks hingga menampilkan hasil klasifikasinya. Alur proses sistem klasifikasi teks dapat dilihat pada gambar 4.5.



Gambar 4.5. Diagram Aktivitas Melakukan Klasifikasi Teks



## 4.2. Desain Halaman Web *Heat Map*

Desain halaman web *heat map* hasil dari pemetaan data Twitter oleh aplikasi OpenGeo Suite meliputi desain basis data, diagram *use case*, diagram sekuen, dan diagram aktivitas

### 4.2.1. Desain Basis Data Halaman Web *Heat Map*

Basis data halaman web *heat map* mempunyai empat tabel dengan atribut yang sama yaitu 'id', 'date', 'latitude', 'longidute', dan 'text'. Keempat tabel tersebut secara berurutan dinamakan 'telkomposdes' untuk periode Desember dengan data sentimen positif, 'telkomnegdes' untuk periode Desember dengan data sentimen negatif, 'telkomposapr' untuk periode April dengan data sentimen positif, dan 'telkomnegapr' untuk periode April dengan data sentimen negatif. Ini bertujuan agar dapat menampilkan lapisan pemetaan yang diinginkan sehingga mengetahui rangkaian waktu (*time series*) dan perbedaan dominasi sentimen positif dan negatif. Basis data yang digunakan adalah PostGIS. Berikut adalah desain tabel halaman web *heat map* seperti terlihat tabel 4.3.

**Tabel 4.3. Desain Basis Data untuk Halaman Web *Heat Map***

telkomposdes		telkomnegdes	
PK	<u>id</u>	PK	<u>id</u>
	date latitude longitude text		date latitude longitude text

telkomposapr		telkomnegapr	
PK	<u>id</u>	PK	<u>id</u>
	date latitude longitude text		date latitude longitude text

Untuk implementasinya, basis data yang digunakan adalah PostGIS. Basis data ini dipilih karena telah sinkron dengan aplikasi untuk pemetaan yaitu GeoServer dan GeoExplorer sehingga data dapat langsung di impor ke aplikasi pemetaan tersebut. Pada saat akan melakukan pemetaan data, GeoServer membutuhkan sebuah data, GeoServer membutuhkan sebuah data, GeoServer membutuhkan sebuah data titik yang dari data masukan. Oleh karena itu ada *query* yang membuat atribut-atribut koordinat yaitu 'latitude' dan 'longitide' dijadikan satu. Atribut baru tersebut dinamakan 'geom'. Seperti terlihat tabel 4.4.

**Tabel 4.4. Desain Basis Data untuk Pemetaan Data**

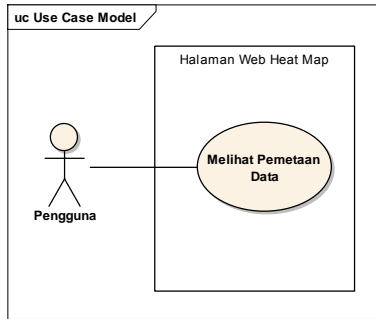
telkomposdes		telkomnegdes	
PK	<u>id</u>	PK	<u>id</u>
	date		date
	geom		geom
	text		text

telkomposapr		telkomnegapr	
PK	<u>id</u>	PK	<u>id</u>
	date		date
	geom		geom
	text		text

#### 4.2.2. Diagram Use Case Halaman Web *Heat Map*

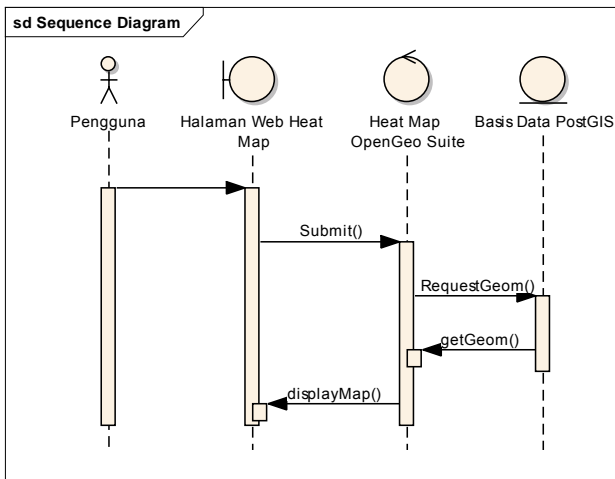
Dalam menggunakan halaman web *heat map* ini, aktor yang terlibat hanya pengguna. Kebutuhan fungsional dari halaman web *heat map* ini adalah dapat melihat pemetaan data yang telah dilakukan oleh aplikasi OpenGeo Suite. Seperti terlihat gambar 4.6



Gambar 4.6. Diagram Use Case Halaman Web *Heat Map*

#### 4.2.3. Diagram Sekuen Melihat Pemetaan Data

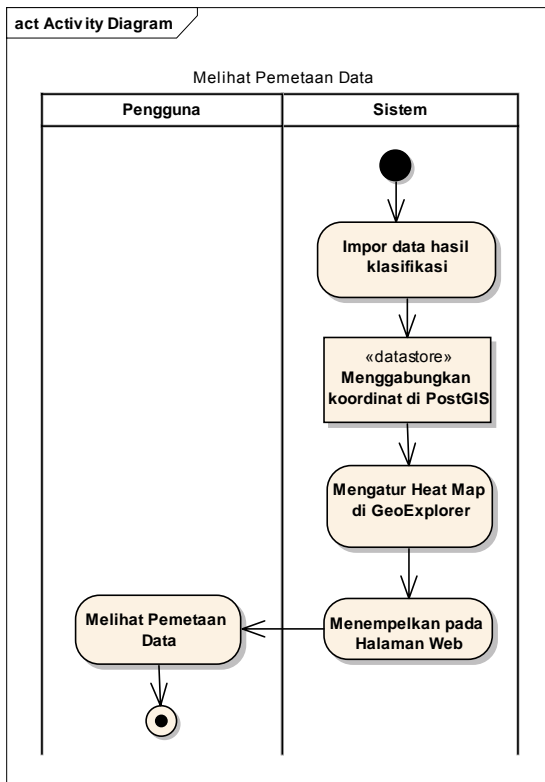
Pada halaman web *heat map* pengguna hanya dapat melakukan melihat pemetaan data, maka diagram sekuen juga hanya melihat pemetaan data. Seperti terlihat gambar 4.7 diagram sekuen dari melihat pemetaan data.



Gambar 4.7. Diagram Sekuen Melihat Pemetaan Data

#### 4.2.4. Diagram Aktivitas Melihat Pemetaan Data

Diagram aktivitas melihat pemetaan data diawali dari impor data hasil klasifikasi sistem pengklasifikasi teks *Naïve Bayes*. Kemudian data diolah agar dapat dipetakan melalui basis data PostGIS. Setelah itu ada pengaturan *heat map* di GeoExplorer yang hasilnya diambil untuk ditempelkan pada halaman web. Sehingga pengguna dapat melihatnya pada halaman web seperti terlihat gambar 4.8



Gambar 4.8. Diagram Aktivitas Melihat Pemetaan Data

*Halaman ini sengaja dikosongkan*

## **BAB V**

### **IMPLEMENTASI APLIKASI**

Bab implementasi aplikasi ini menjelaskan implementasi desain aplikasi pengklasifikasi teks. Implementasi desain aplikasi pengklasifikasi teks ini menggunakan bahasa pemrograman PHP. Selain itu bab ini juga menjelaskan implementasi desain halaman web *heat map* OpenGeo Suite untuk memetakan data hasil klasifikasi teks dalam sebuah halaman *web*.

#### **5.1. Implementasi Aplikasi Pengklasifikasi Teks**

Bagian ini menjelaskan proses pembuatan aplikasi klasifikasi teks berbasis *Naïve Bayes*. Aplikasi ini bertujuan untuk mengklasifikasikan suatu *tweet* ke dalam kelas sentimen positif atau negatif. Dengan begitu proses klasifikasinya berlangsung secara otomatis. Proses klasifikasi teks dimulai dengan memasukkan file data berformat csv ke dalam aplikasi. File tersebut berisikan *tweets* yang hendak diklasifikasikan namun belum diketahui apakah termasuk opini positif atau negatif. Berikut teknis pembuatan aplikasi klasifikasi teks:

##### **5.1.1. Pembuatan Basis Data di MySQL**

Berdasarkan bagian 4.1.1 desain basis data pengklasifikasi teks, implementasi metode *Naïve Bayes* membutuhkan basis data MySQL untuk menampung seleksi fitur yang berisikan kumpulan kata, kelas sentimen, serta frekuensi kemunculan. Total fitur yang diambil untuk dijadikan sebagai data pelatihan adalah 336 kata. Tiap fitur memiliki dua nilai frekuensi yaitu positif dan negatif. Masing-masing kelas harus mempunyai nilai untuk sebuah fitur. Sehingga misalkan sebuah kata tersebut adalah ‘bagus’, maka pemberian nilai positif haruslah lebih besar dari nilai negatif. Karena kata tersebut mencerminkan klasifikasi yang positif. Begitu pula sebaliknya jika kata yang muncul adalah kata negatif seperti ‘jelek’, maka

pada sentimen kelas negatif nilainya harus lebih besar daripada sentimen kelas positif. Untuk penentuan besaran nilai bergantung dari keseringan muncul dalam sebuah himpunan data pelatihan.

Berdasarkan tujuan dari pengerjaan tugas akhir ini yaitu memberikan manfaat CRM pada perusahaan Telkomsel dari sisi operasional, maka ada pengkategorian operasional CRM yang menghasilkan fitur-fitur pilihan. Dengan memperhatikan sisi operasional CRM untuk perusahaan penyedia layanan komunikasi ini, ada empat kategori yang dihasilkan diantaranya kategori kepuasan, kategori keluhan, kategori loyalitas, dan kategori respons. Kemudian fitur-fitur yang dipilih dari dokumen pelatihan harus bersesuaian dengan keempat kategori tersebut. Sehingga hasilnya seperti terlihat pada tabel 5.1

**Tabel 5.1. Kategori Fitur Operasional CRM**

Kepuasan	Keluhan	Loyalitas	Respons
asik	lelah	selingkuhannya	rame
thanks	lambat	migrasi	menurun
trust	songong	pacar	jomblo
aman	jelek	setia	antri
manteep	gagal	bebulik	tertarik
seneng	bacut	prhatian	gak
maacih	dikibuli	seperhatian	tidak
tengkyu	nyesek	putus	huft
makasih	lemah	ganti	kurang
bagus	ilang	mendingan	astaga

Fitur tersebut diperoleh dari data dokumen pelatihan dengan cara manual. Proses pemilihan fitur dilakukan dengan berpedoman pada empat kategori operasional CRM tersebut. Fitur-fitur ini kemudian dimasukkan ke dalam basis data MySQL. Seperti yang ditunjukkan oleh tabel 5.2.

Tabel 5.2. Basis Data Fitur Kata

idteks	teks	Pscore	Nscore	idteks	teks	Pscore	Nscore
1460001	rame	3	4	1460031	lelet	15	16
1460002	asik	1	0	1460032	makasih	7	6
1460003	thanks	5	4	1460033	tdk	4	5
1460004	kuat	4	3	1460034	ganteng	1	0
1460005	kurang	3	4	1460035	emosi	2	3
1460006	dikibuli	0	1	1460036	cinta	1	0
1460007	gak	58	59	1460037	lancak	1	0
1460008	gaada	0	1	1460038	aduh	1	2
1460009	kek	4	5	1460039	sabar	0	1
1460010	mulu	27	28	1460040	belum	2	3
1460011	manteep	1	0	1460041	kesel	6	7
1460012	songong	0	1	1460042	eror	1	2
1460013	gabisa	3	4	1460043	thankuuu	1	0
1460014	mantep	1	0	1460044	nyangkut	0	1
1460015	jauuuuh	0	1	1460045	ngeseli	2	3
1460016	jelek	18	19	1460046	iyalah	1	0
1460017	aman	4	3	1460047	dasar	0	1
1460018	ilang	22	23	1460048	m*h*l	0	1
1460019	gagal	9	10	1460049	selingkuhannya	0	1
1460020	tidak	17	18	1460050	tinggal	0	1
1460021	lemah	2	3	1460051	benci	1	2
1460022	susah	15	16	1460052	wedus	0	1
1460023	lemot	51	52	1460053	tahayul	1	2
1460024	cukkk	0	1	1460054	ngadat	4	5
1460025	lelah	1	2	1460055	huft	2	3



Pengkategorian fitur ini juga berlaku untuk fitur *emoticon*. Seleksi fitur yang dilakukan tidak hanya berupa kata-kata sentimen positif atau negatif, namun juga *emot* yang menandakan sentimen kelas positif atau negatif. Misalkan ada *emot* seperti ‘-\_-’ atau ‘:D’. Kedua *emot* tersebut memiliki nilai yang berbeda untuk sentimen kelasnya. Untuk *emot* ‘-\_-’ cenderung mengartikan sentimen kelas negatif, sedangkan *emot* ‘:D’ artinya sentimen kelas positif (lihat tabel 5.3)

**Tabel 5.3. Basis Data Fitur *Emoticons***

idteks	teks	Pscore	Nscore	idteks	teks	Pscore	Nscore
1460313	=))	2	1	1460313	=))	2	1
1460314	:/	0	1	1460314	:/	0	1
1460315	:(	19	20	1460315	:(	19	20
1460316	:D	2	1	1460316	:D	2	1
1460317	x_x	2	3	1460317	x_x	2	3
1460318	-____-	0	1	1460318	-____-	0	1
1460319	:)	5	4	1460319	:)	5	4
1460320	:-(	0	1	1460320	:-(	0	1
1460321	:*	2	1	1460321	:*	2	1
1460322	-____-	0	1	1460322	-____-	0	1
1460323	-__-	0	1	1460323	-__-	0	1
1460324	-____-	1	2	1460324	-____-	1	2
1460325	:v	1	0	1460325	:v	1	0
1460326	-____-	0	1	1460326	-____-	0	1
1460327	B-)	1	0	1460327	B-)	1	0
1460328	=	1	2	1460328	=	1	2
1460329	:3	1	0	1460329	:3	1	0
1460330	x_x	3	4	1460330	x_x	3	4
1460331	(y)	1	0	1460331	(y)	1	0
1460332	(n)	0	1	1460332	(n)	0	1
1460333	!!!!	6	7	1460333	!!!!	6	7
1460334	LELET	0	1	1460334	LELET	0	1
1460335	NGEYEL	0	1	1460335	NGEYEL	0	1
1460336	(:)	2	3	1460336	(:)	2	3

## 5.1.2. Implementasi Fungsi Mengunggah Data Dokumen

Sesuai dengan yang dipaparkan di bagian 4.1.2 diagram *use case* pengklasifikasi teks, ada dua kebutuhan fungsional yang salah satunya merupakan kebutuhan mendasar dalam menerapkan klasifikasi teks yakni aplikasi harus dapat mengunggah data dokumen pelatihan dengan ekstensi dan aturan *header* tertentu seperti di bagian 4.1.4.1 diagram aktivitas mengunggah data dokumen. Naskah kode dalam mengimplementasikan fungsi mengunggah data dokumen sebagai berikut ditunjukkan kode 5.1.

<i>main.php</i>	
No	Kode
1	<code>&lt;?php</code>
2	<code>  if(isset(\$_POST['submit'])){</code>
3	<code>    include('class/csv_processing.php');</code>
4	<code>    include('class/db/config.php');</code>
5	<code>    \$table_name = 'tweet';</code>
6	<code>    \$csv = new Csv_processing();</code>
7	<code>    \$allowedExts = array("csv");</code>
8	<code>    \$tmp = explode(".", \$_FILES["file"]["name"]);</code>
9	<code>    \$extension = end(\$tmp);</code>
10	<code>    if (in_array(\$extension, \$allowedExts)){</code>
11	<code>      if (\$_FILES["file"]["error"] &gt; 0){</code>
12	<code>        echo "Error: " . \$_FILES["file"]["error"] .</code>
13	<code>"&lt;br&gt;";</code>
14	<code>      } else {</code>
15	<code>        \$header = array('id','date','text',</code>
16	<code>        'longitude','latitude','sentiment');</code>
17	<code>        \$data = \$csv-</code>
18	<code>&gt;convert_csv_to_array(\$_FILES["file"]["tmp_name"]);</code>
19	<code>        \$header_data = "";</code>
20	<code>        foreach(\$header as \$row){</code>
21	<code>          \$header_data .= \$row.",";</code>
22	<code>        }</code>
23	<code>        \$header_data = substr(\$header_data, 0,</code>
24	<code>        strlen(\$header_data)-1);</code>
25	<code>        \$csv-&gt;start_table();</code>
26	<code>        \$csv-&gt;start_row();</code>
27	<code>        \$csv-&gt;set_header(\$header);</code>
28	<code>        \$csv-&gt;end_row();</code>
29	<code>        foreach(\$data as \$key=&gt;\$val){</code>

**Kode 5.1. Query untuk Mengunggah Data Dokumen**

<i>main.php</i>	
No	Kode
30	<code>\$csv-&gt;start_row();</code>
31	<code>\$tweet_data = "";</code>
32	<code>foreach(\$val as \$key2=&gt;\$val2){</code>
33	<code>  \$csv-&gt;set_td(\$val2);</code>
34	<code>  if(\$key2 == 'id') \$tweet_data .=</code>
35	<code>  \$val2.",";</code>
36	<code>  else \$tweet_data .= "'".\$val2."',";</code>
37	<code>  }</code>
38	<code>  \$tweet_data =</code>
39	<code>  substr(\$tweet_data, 0, strlen(\$tweet_data)-1);</code>
40	<code>  mysql_query("INSERT INTO ".\$table_name."</code>
41	<code>  ( ".\$header_data." ) VALUES ( ".\$tweet_data." )");</code>
42	<code>  \$csv-&gt;end_row();</code>
43	<code>  }</code>
44	<code>\$csv-&gt;end_table();</code>
45	<code>}</code>
46	<code>} else {</code>
47	<code>  echo 'file harus .csv';</code>
48	<code>}</code>
49	<code>} else{</code>
50	<code>  \$form_link = "http://" .</code>
51	<code>  \$_SERVER['HTTP_HOST'] . \$_SERVER['REQUEST_URI'] . ' ';</code>
52	<code>  include('views/main_view.php');</code>
53	<code>}</code>
54	<code>?&gt;</code>

**Kode 5.1. Query untuk Mengunggah Data Dokumen (lanjutan)**

### 5.1.3. Implementasi Pengklasifikasi Teks

*Naïve Bayes* adalah metode pembelajaran yang diawasi sehingga membutuhkan contoh data sebelum melakukan klasifikasi. Keberhasilan metode ini sangat bergantung pada pemilihan fitur dari contoh data. Misalkan berdasarkan contoh *tweet* opini negatif sebelumnya, fitur yang mewakili bahwa *tweet* tersebut termasuk kelas sentimen negatif adalah ‘kecewa’ dan ‘mahal’. Sedangkan untuk *tweet* opini positif dapat diambil fitur ‘love’. Dari hal ini kemudian langkah selanjutnya memasukkan fitur ‘kecewa’, ‘mahal’, dan ‘love’ ke dalam basis data sebagai data pelatihan. Jika terjadi kesalahan dalam pemilihan fitur, maka hasil klasifikasi kemungkinan juga kurang tepat.

Penerapan *Naïve Bayes* di dalam bahasa pemrograman PHP membutuhkan kelas *text\_mining.php* sebagai algoritma *Naïve Bayes*, dan *text\_model.php* sebagai pengambilan jumlah frekuensi per kata di basis data sebagaimana dijelaskan pada bagian 4.1.3.2 diagram sekuen melakukan klasifikasi teks. Masing-masing kelas saling berhubungan dalam membangun perangkat lunak *Naïve Bayes*. Naskah kode *Naïve Bayes* dapat dilihat pada kode 5.2.

<b><i>text_mining.php</i></b>	
No	Kode
1	<?php
2	class Text_mining{
3	function __construct(){
4	require_once('text_model.php');
5	\$this->model = new Text_model();
6	}
7	private \$model;
8	private \$category = 2;
9	function get_sentiment_result(\$teks){
10	\$sum_frekuensi = \$this->model->get_sum_frekuensi();
11	\$all_keywords = \$this->model->get_all_keywords();
12	\$neg_val = 0.5;
13	\$pos_val = 0.5;
14	foreach(\$all_keywords as \$row){
15	if(strpos(\$teks, \$row['teks']) !== false){
16	\$neg_val *=
17	(\$row['Nscore']+1)/\$sum_keywords['total'];
18	\$pos_val *=
19	(\$row['Pscore']+1)/sum_keywords['total'];
20	}
21	}
22	if(\$neg_val > \$pos_val) \$sentiment = 'negative';
23	else \$sentiment = 'positive';
24	\$result = array(
25	'indexpos' => \$pos_val,
26	'indexneg' => \$neg_val,
27	'sentiment' => \$sentiment
28	);
29	var_dump(\$result);
30	return \$result;
31	}
32	}
33	?>

**Kode 5.2. Query Metode Naïve Bayes**

Sebagaimana konsep *Naïve Bayes* yang dijelaskan pada sub bab 2.3.1. Metode *Naïve Bayes*, penerapan algoritma *text\_mining.php* diawali dengan inisiasi hasil bagi jumlah kelas yaitu positif dan negatif. Karena ada dua kelas maka nilainya 0,5. Kemudian program mencari kata yang sama dengan yang ada di basis data. Jika menemukan kata yang sama, maka nilai inisiasi dikalikan dengan tiap-tiap frekuensi kata yang muncul dalam data sampel. Itu diterapkan pada kelas positif dan negatif, lalu penentuan kelas melihat dari hasil perhitungan yang memiliki nilai yang lebih besar. Berikut adalah pembahasan lebih merinci dari kelas *text\_mining.php*.

Naskah kode 5.3 berfungsi untuk membangun hubungan basis data dengan kelas *text\_model.php*. Hal ini sama dengan desain yang dijelaskan pada bagian 4.1.3.2 diagram sekuen melakukan klasifikasi teks, mengenai fungsi *text\_model.php* sebagai penghubung dengan basis data. Selain itu juga untuk mendeklarasikan variable *model* dan *category* yang diberi nilai dua sebagaimana jumlah kelas yang ada. Karena prosesnya mengambil fitur yang muncul dari data contoh secara langsung.

<i>text_mining.php</i>	
No	Kode
1	<?php
2	class Text_mining{
3	function __construct() {
4	require_once('text_model.php');
5	\$this->model = new Text_model();
6	}
7	private \$model;
8	private \$category = 2;

**Kode 5.3. Query Inisiasi Variabel *Naïve Bayes***

Fungsi *get\_sentiment results(\$teks)* pada naskah kode 5.4 berguna untuk menjumlahkan keseluruhan frekuensi kata yang ada di basis data. Kemudian mencari kata positif di contoh data apakah ada yang sama dengan di basis data. Setiap ada kata positif yang

sama, maka nilai tersebut disimpan untuk kemudian dihitung pada kode berikutnya.

<i>text_mining.php</i>	
No	Kode
9	<code>function get_sentiment_result(\$teks){</code>
10	<code>  \$sum_frekuensi = \$this-&gt;model-&gt;get_sum_frekuensi();</code>
11	<code>  \$all_keywords = \$this-&gt;model-&gt;get_all_keywords();</code>
12	<code>  \$neg_val = 0.5;</code>
13	<code>  \$pos_val = 0.5;</code>

**Kode 5.4. Query Fungsi *get\_sentiment\_result***

Naskah kode 5.5 menunjukkan perhitungan *Naïve Bayes*. Dimana *\$neg\_val* atau *\$pos\_val* dikalikan dengan hasil bagi jumlah frekuensi fitur yang ditemukan dengan jumlah total fitur yang ada. Kemudian langkah selanjutnya membandingkan hasil antara *\$neg\_val* dan *\$pos\_val*. Jika nilai *\$neg\_val* lebih besar dari *\$pos\_val*, maka keluarannya negatif. Selain itu jika nilai *\$pos\_val* lebih besar dari *\$neg\_val*, maka keluarannya positif.

<i>text_mining.php</i>	
No	Kode
14	<code>  foreach(\$all_keywords as \$row){</code>
15	<code>  if(strpos(\$teks, \$row['teks']) !== false){</code>
16	<code>    \$neg_val *=</code>
17	<code>      (\$row['Nscore']+1)/\$sum_keywords['total'];</code>
18	<code>    \$pos_val *=</code>
19	<code>      (\$row['Pscore']+1)/sum_keywords['total'];</code>
20	<code>  }</code>
21	<code>}</code>
22	<code>if(\$neg_val &gt; \$pos_val) \$sentiment = 'negative';</code>
23	<code>else \$sentiment = 'positive';</code>

**Kode 5.5. Query Perhitungan Fitur *Naïve Bayes***

Naskah kode 5.6 berguna untuk menampilkan hasil perhitungan antara *\$pos\_val* dan *\$neg\_val*. Masing-masing diterjemahkan dalam 'indexpos' untuk *\$pos\_val* dan 'indexneg'

untuk  $\$neg\_val$ , lalu hasilnya berupa kelas positif atau negatif ditampilkan melalui 'sentiment'.

<i>text_mining.php</i>	
No	Kode
24	<code>\$result = array(</code>
25	<code>  'indexpos' =&gt; \$pos_val,</code>
26	<code>  'indexneg' =&gt; \$neg_val,</code>
27	<code>  'sentiment' =&gt; \$sentiment</code>
28	<code>);</code>
29	<code>var_dump(\$result);</code>
30	<code>return \$result;</code>
31	<code>  }</code>
32	<code>}</code>
33	<code>?&gt;</code>

**Kode 5.6. Query Hasil Metode Naïve Bayes**

Selain kelas *text\_mining.php*, ada juga kelas lain bernama *text\_model.php* seperti kode 5.7. Kelas *text\_mining.php* membutuhkan kelas *text\_model.php* sebagai pemanggil fitur-fitur di basis data MySQL. Perhitungan NBC menggunakan fitur-fitur yang ada di basis data MySQL dengan tabel *keyword*. Berikut adalah naskah kode *text\_model.php* secara lengkap:

<i>text_model.php</i>	
No	Kode
1	<code>&lt;?php</code>
2	<code>  class Text_model{</code>
3	<code>    function __construct(){</code>
4	<code>      include('db/config.php');</code>
5	<code>    }</code>
6	<code>    function get_all_keywords(){</code>
7	<code>      \$query = "SELECT * FROM keyword";</code>
8	<code>      \$result = mysql_query(\$query);</code>
9	<code>      \$arr = array();</code>
10	<code>      while(\$row = mysql_fetch_array(\$result)){</code>
11	<code>        \$arr[] = \$row;</code>
12	<code>      }</code>
13	<code>      return \$arr;</code>

**Kode 5.7. Query text\_model.php**

<b><i>text_model.php</i></b>	
No	Kode
14	}
15	function get_total_sum(){
16	\$cquery = "SELECT COUNT(*) as count FROM
17	keyword";echo \$cquery;
18	\$cresult = mysql_query(\$cquery);echo 'num:
19	' .mysql_num_rows(\$cresult);
20	\$c = 0;
21	while(\$row = mysql_fetch_array(\$cresult)){
22	\$c = \$row['count'];
23	}
24	\$pquery = "SELECT SUM(Pscore) as total FROM
25	keyword";
26	\$presult = mysql_query(\$pquery);
27	\$psum = 0;
28	while(\$row = mysql_fetch_array(\$presult)){
29	\$psum = \$row['total'] + \$c;
30	}
31	\$nquery = "SELECT SUM(Nscore) as total FROM
32	keyword";
33	\$nresult = mysql_query(\$nquery);
34	\$nsum = 0;
35	while(\$row = mysql_fetch_array(\$nresult)){
36	\$nsum = \$row['total'] + \$c;
37	}
38	return array(
39	'negative' => \$nsum,
40	'positive' => \$psum,
41	'total' => (intval(\$psum) + intval(\$nsum))
42	);
43	}

**Kode 5.7. Query *text\_model.php* (lanjutan)**

Naskah kode 5.8 menunjukkan *text\_model.php* yang diawali dengan memanggil kelas *config.php*. Di dalamnya ada pengaturan koneksi ke basis data yang berisikan himpunan kata positif dan negatif. Fungsi dari kelas *text\_model.php* untuk menghubungkan proses klasifikasi teks dengan fitur yang tersimpan di basis data.



<i>text_model.php</i>	
No	Kode
1	<?php
2	class Text_model{
3	function __construct(){
4	include('db/config.php');
5	}

**Kode 5.8. Query untuk Koneksi ke Basis Data**

Naskah kode 5.9 menunjukkan fungsi *get\_all\_keywords()* untuk mendapatkan semua kata dari tabel 'keyword'. Kelas *text\_model.php* membuat kelas *text\_mining.php* membandingkan kata-kata dari data contoh dengan yang ada di basis data. Kemudian kata-kata diubah menjadi *array* dengan *mysql\_fetch\_array*.

<i>text_model.php</i>	
No	Kode
6	function get_all_keywords(){
7	\$query = "SELECT * FROM keyword";
8	\$result = mysql_query(\$query);
9	\$arr = array();
10	while(\$row = mysql_fetch_array(\$result)){
11	\$arr[] = \$row;
12	}
13	return \$arr;
14	}

**Kode 5.9. Query Fungsi *get\_all\_keywords***

Naskah kode 5.10 menunjukkan fungsi *get\_total\_sum()* untuk menghitung total frekuensi fitur yang ada pada tabel 'keyword'. Fungsi ini menggunakan *query* "SELECT COUNT(\*) as count FROM keyword". Total frekuensi fitur dituliskan dengan variabel *\$query*. Selain itu juga perlu variabel *\$c* dengan inisialisasi 0 yang akan terus bertambah seiring hasil penjumlahan dari frekuensi fitur.

<i>text_model.php</i>	
No	Kode
15	<code>function get_total_sum(){</code>
16	<code>    \$query = "SELECT COUNT(*) as count FROM</code>
17	<code>keyword";echo \$query;</code>
18	<code>    \$result = mysql_query(\$query);echo 'num:</code>
19	<code>    '.mysql_num_rows(\$result);</code>
20	<code>    \$c = 0;</code>
21	<code>    while(\$row = mysql_fetch_array(\$result)){</code>
22	<code>        \$c = \$row['count'];</code>
23	<code>    }</code>

**Kode 5.10. Query untuk Total Frekuensi Fitur**

Naskah kode 5.11 menunjukkan fungsi `get_total_sum()` dan *query* “SELECT SUM(Pscore) as total FROM keyword” yang berfungsi untuk menjumlahkan frekuensi fitur yang muncul dari data contoh. Total frekuensi fitur positif juga di inisialisasi dengan nilai 0. Nilai ini akan bertambah seiring bertambahnya hasil penjumlahan frekuensi fitur positif yang ditemukan pada data contoh. Sedangkan untuk fitur negatif menggunakan *query* “SELECT SUM(Nscore) as total FROM keyword”. Sama halnya dengan fungsi untuk perhitungan fitur positif, fungsi perhitungan fitur negatif juga memulai dengan inisialisasi nilai 0.

<i>text_model.php</i>	
No	Kode
24	<code>    \$query = "SELECT SUM(Pscore) as total FROM</code>
25	<code>keyword";</code>
26	<code>    \$result = mysql_query(\$query);</code>
27	<code>    \$psum = 0;</code>
28	<code>    while(\$row = mysql_fetch_array(\$result)){</code>
29	<code>        \$psum = \$row['total'] + \$c;</code>
30	<code>    }</code>
31	<code>    \$query = "SELECT SUM(Nscore) as total FROM</code>
32	<code>keyword";</code>
33	<code>    \$result = mysql_query(\$query);</code>
34	<code>    \$nsum = 0;</code>
35	<code>    while(\$row = mysql_fetch_array(\$result)){</code>
36	<code>        \$nsum = \$row['total'] + \$c;</code>

**Kode 5.11. Query Fungsi `get_total_sum`**

<i>text_model.php</i>	
No	Kode
37	}
38	return array(
39	'negative' => \$nsum,
40	'positive' => \$psum,
41	'total' => (intval(\$psum) + intval(\$nsum))
42	);
43	}

**Kode 5.11. Query Fungsi *get\_total\_sum* (lanjutan)**

Secara keseluruhan tahapan sesuai dengan bagian 4.1.4.2 diagram aktivitas melakukan klasifikasi teks. Awalnya program klasifikasi teks berbasis *Naïve Bayes* membuat proses klasifikasi berjalan secara otomatis. Data teks Twitter yang belum diketahui kelasnya, kemudian dimasukkan ke dalam program ini untuk dapat diketahui kelasnya secara langsung. Program ini hanya memperkenankan data file ekstensi csv. File yang dimasukkan hanya terdiri dari baris-baris teks tanpa ada atribut lainnya. Tampilan program klasifikasi *Naïve Bayes* ditunjukkan pada gambar 5.3.

**Gambar 5.1. Program Klasifikasi Teks Naive Bayes**

Setelah file dimasukkan dan diklik *submit*, maka program melakukan proses sebagaimana logika algoritma *Naïve Bayes* yang sudah dijelaskan pada sub bab sebelumnya. Keluaran yang dihasilkan adalah besaran indekspos dan indeksneg, serta kesimpulan kelas sentimen yang diambil dari nilai indekspos atau indeksneg yang lebih besar. Contohnya adalah beberapa data yang diperoleh pada tanggal 10 April 2014 seperti terlihat tabel 5.4.

Tabel 5.4. Data Contoh Sebelum diklasifikasi

teks	sentiment
<i>masbudilo aku pakek 2 pin bud! Kan pakek telkom aku tapi gang ku ini gapernah ada sinyal telkomsel -_-</i>	Null
<i>Telkomsel Sucks!!!</i>	Null
<i>Aku pake axis. Lama banget sumvah -_- RT shintaeun: widiainia pake telkomsel bendera merah burung garuda!!!</i>	Null

Setelah ditekan 'submit', maka program memulai proses klasifikasi teks. Tabel 5.5 menunjukkan hasil pemrosesan program klasifikasi teks

Tabel 5.5. Data Contoh Hasil Klasifikasi

teks	indekspos	indeksneg	sentiment
<i>masbudilo aku pakek 2 pin bud! Kan pakek telkom aku tapi gang ku ini gapernah ada sinyal telkomsel-_-</i>	2.287740570 2193E-7	<b>4.117933026</b> <b>3948E-7</b>	negative
<i>Telkomsel Sucks!!!</i>	0.000320855 61497326	<b>0.000427807</b> <b>48663102</b>	negative
<i>Aku pake axis. Lama banget sumvah -_- RT shintaeun: widiainia pake telkomsel bendera merah burung garuda!!!</i>	<b>7.633957838</b> <b>5929E-10</b>	5.872275260 4561E-10	positive

Oleh karena itu, seluruhnya dimasukkan ke dalam program ini untuk diklasifikasikan ke dalam kelas sentimen positif atau negatif. Kemudian hasil klasifikasi sentimen positif dan negatif dikelompokkan menjadi dua klaster yang berbeda. Hal ini bertujuan untuk memudahkan dalam membuat pemetaan data. Hasil klasifikasi

teks secara langsung masuk ke dalam basis data. Basis data yang telah terisi data contoh hasil klasifikasi dapat dilihat pada tabel 5.6

**Tabel 5.6. Hasil Klasifikasi di dalam Basis Data**

id	date	text	longitude	latitude	sentiment
14060001	Thu Apr 10 16:05:22 +0000 2014	masbudilo aku pakek 2 pin bud! Kan pakek telkom aku tapi gang ku ini gapernah ada sinyal telkomsel-_-	3.2502055	98.5395362	negative
14060002	Thu Apr 10 16:37:30 +0000 2014	Telkomsel Sucks!!!	-4.04767	122.46685	negative
14060003	Thu Apr 10 16:54:28 +0000 2014	Aku pake axis. Lama banget sumvah -_- RT shintaean: widiainia pake telkomsel bendera merah burung garuda!!!	-6.2539583	106.8498934	positive

## 5.2. Implementasi Halaman Web *Heat Map*

Proses pemetaan data menggunakan OpenGeo Suite 3.1. Aplikasi ini terdiri dari komponen-komponen untuk pemetaan yang saling terintegrasi. Data yang telah melalui klasifikasi teks dimasukkan ke salah satu komponen OpenGeo Suite yang berfungsi sebagai basis data yaitu PostGIS. Basis data khusus pemetaan ini memiliki keunggulan mampu menggabungkan dua titik (bujur dan lintang) menjadi sebuah titik koordinat (*geometry*) sehingga memungkinkan untuk diproses pada komponen selanjutnya. Maka selanjutnya data yang telah disiapkan dimasukkan ke aplikasi PostGIS.

### 5.2.1. Pembuatan Basis Data di PostGIS

Data hasil klasifikasi diimpor ke dalam PostGIS melalui beberapa tahap *query*. Sebelumnya file data hasil klasifikasi berformat teks dipersiapkan pada folder tertentu. Kemudian pada PostGIS diketikkan *query* untuk membuat tabel yang disesuaikan dengan atribut data. Untuk naskah kode membuat tabel baru dapat dilihat pada kode 5.12

<i>membuat tabel.sql</i>	
No	Kode
1	<code>CREATE TABLE telkomposdes (</code>
2	<code>  id INTEGER PRIMARY KEY,</code>
3	<code>  text VARCHAR(500),</code>
4	<code>  latitude FLOAT8,</code>
5	<code>  longitude FLOAT8,</code>
6	<code>  date VARCHAR(300));</code>

**Kode 5.12. Query membuat\_tabel.sql**

Sebagaimana di bagian 4.2.1 desain basis data halaman web *heat map*, untuk pemetaan ini, ada tiga atribut yang akan ditampilkan yaitu id, text, dan date. Sedangkan atribut latitude dan longitude akan melebur dan berbentuk titik. Setelah menjadi tabel kosong dengan lima kolom, lalu file dengan lima kolom juga diimpor ke dalam PostGIS. Query untuk membuat tabel baru dinamakan *telkomposdes* dengan query *membuat\_tabel.sql* dapat dilihat pada naskah kode 5.13

<i>impor_data.sql</i>	
No	Kode
1	<code>COPY telkomposdes FROM '/telkomposdes.txt' WITH (</code>
2	<code>  FORMAT csv,</code>
3	<code>  DELIMITER E'\t',</code>
4	<code>  QUOTE '*',</code>
5	<code>  HEADER false,</code>
6	<code>  ENCODING 'UTF8'</code>
7	<code>);</code>

**Kode 5.13. Query impor\_data.sql**

Setelah data diimpor, tabel *telkomposdes* terisi dua titik koordinat. Namun GeoServer tidak dapat membaca dua titik tersebut. Jadi harus digabungkan menjadi sebuah kolom koordinat. Data yang diimpor melalui *impor\_data.sql* yang mengacu pada lokasi file tertentu. Lokasi ini bisa diatur sesuai keberadaan file. Dalam hal ini ada lima kolom yang dimasukkan tabel sesuai jumlah kolom yang sudah dibuat sebelumnya seperti gambar 5.2.

	id [PK] integer	date character var	latitude double precis	longitude double precis	text character varying(500)
1	6614001	Sun Dec 22	-7.3252342	110.3212587	Telkomsel tumben sinyal
2	6614002	Sun Dec 22	-6.32661	107.31049	Manteep sinyal telkoms
3	6614003	Sun Dec 22	-6.94949	109.81772	Di mana? Di kerawang do
4	6614004	Sun Dec 22	-7.17227	107.57348	Ga ada sms di, sinyalnya
5	6614005	Sun Dec 22	-0.4373489	116.9890137	Telkomsel g ada kakak a
6	6614006	Sun Dec 22	-6.24305	106.90027	Telkomsel yah susah deh
7	6614007	Sun Dec 22	-3.79697224	102.2665835	Bengkulu nih sinyal tel
8	6614008	Mon Dec 23	-5.43504	105.26328	Pake telkomsel biar gan
9	6614009	Mon Dec 23	-6.3212606	106.7278899	Ini gua cinta banget ka
10	6614010	Mon Dec 23	-4.86074	104.57232	Ketika tablet pake tiga

Gambar 5.2. Data Hasil Impor ke PostGIS

Kode 5.14 menunjukkan *query* untuk membuat tabel baru yang mengakomodasi kolom penggabungan koordinat. Tabel baru ini diberi nama *telkomselposdes*. Data yang sudah dimasukkan di tabel *telkomposdes* secara otomatis langsung pindah ke tabel *telkomselposdes*.

<i>membuat_tabel_koordinat.sql</i>	
No	Kode
1	
2	<code>SELECT Count(*) FROM telkomposdes;</code>
3	<code>CREATE TABLE telkomselposdes AS</code>
4	<code>SELECT id AS id,</code>
5	<code>text AS text,</code>
6	<code>date AS date,</code>
7	<code>ST_SetSRID(ST_MakePoint(latitude, longitude),</code>
8	<code>4326)::Geometry(Point,4326) AS geom</code>
9	<code>FROM telkomposdes;</code>
10	

Kode 5.14. *Query membuat\_tabel\_koordinat.sql*

## 5.2.2. Penggabungan Atribut Koordinat

Tahapan ini sesuai dengan bagian 4.2.4 diagram aktivitas melihat pemetaan data, setelah data diimpor proses melakukan

penggabungan atribut koordinat. Kode 5.15 menunjukkan *query* untuk menggabungkan koordinat longitude dan latitude menjadi sebuah titik. Tabel yang digunakan adalah tabel yang baru, *telkomselposdes*. Dengan begitu komponen GeoServer dapat membaca titik koordinat tersebut dan dapat memetakannya.

<i>penggabungan_koordinat.sql</i>	
No	Kode
1	
2	<code>ALTER TABLE telkomselposdes ADD PRIMARY KEY (id);</code>
3	<code>CREATE INDEX telkomselposdes_gix ON telkomselposdes</code>
4	<code>USING GIST (geom);</code>
5	

**Kode 5.15.** *Query penggabungan\_koordinat.sql*

Setelah data digabungkan, tabel *telkomselposdes* terisi satu titik koordinat seperti terlihat gambar 5.3. Satu titik koordinat ini memungkinkan GeoServer untuk membaca letak seksama di koordinat bola dunia.

	<b>id [PK] integer</b>	<b>text character varying(500)</b>	<b>date character va</b>	<b>geom geometry(Point,4326)</b>
<b>1</b>	6614001	Telkomsel tumben sinyalnya ku	Sun Dec 22	0101000020E61000008
<b>2</b>	6614002	Manteep sinyal telkomsel no	Sun Dec 22	0101000020E61000000
<b>3</b>	6614003	Di mana? Di kerawang doang ka	Sun Dec 22	0101000020E61000003
<b>4</b>	6614004	Ga ada sms di,sinyalnya jelek	Sun Dec 22	0101000020E61000003
<b>5</b>	6614005	Telkomsel g ada kakak alhamdu	Sun Dec 22	0101000020E6100000F
<b>6</b>	6614006	Telkomsel yah susah deh inter	Sun Dec 22	0101000020E61000007
<b>7</b>	6614007	Bengkulu nih sinyal telkomsel	Sun Dec 22	0101000020E6100000C
<b>8</b>	6614008	Pake telkomsel biar ganteng j	Mon Dec 23	0101000020E61000003
<b>9</b>	6614009	Ini gua cinta banget kali sm	Mon Dec 23	0101000020E6100000A
<b>10</b>	6614010	Ketika tablet pake tiga nyamp	Mon Dec 23	0101000020E61000003

**Gambar 5.3.** Data Hasil Penggabungan Koordinat



### 5.2.3. Pengaturan *Heat Map* di GeoServer

Data yang sudah berada di PostGIS harus diimpor ke dalam aplikasi GeoServer. Karena sudah saling terintegrasi, GeoServer menyediakan menu untuk impor data dari PostGIS. Setelah berhasil diimpor, data secara kealpaan jika dipetakan akan membentuk titik-titik di peta. Karena tidak sesuai yang diharapkan maka perlu mendesain mode *heat map* agar tampilan data sesuai dengan tujuan Tugas Akhir ini.

*Heat map* sendiri merupakan konsep dari *clustering* yang menggunakan metode *max linkage* atau *complete linkage*. Metode ini menghubungkan antar titik kemiripan yang saling berjauhan. Sehingga menampilkan peta dengan klaster-klaster yang cenderung mengurangi kemungkinan riuh (*noise*) atau *outlier*.

Di *toolbar* 'data lalu style' terdapat menu 'add a new style'. Menu ini menyediakan naskah kode kosong untuk membuat mode tampilan sesuai yang diinginkan. Di samping itu, untuk mode *heat map*, oleh OpenGeo Suite 3.1 telah memiliki *Web Map Service* (WMS) yang mana naskah kode telah tersedia di dalamnya. Sehingga hanya perlu mencantumkan layanan WMS di dalam naskah kode baru, maka tampilan *heat map* sudah dapat ditayangkan. Untuk itu, sesuai Relevansi atau Manfaat Tugas Akhir ini, perlu dibedakan antara *heat map* untuk klaster positif dan klaster negatif agar pada saat ditayangkan dapat lebih mudah mengetahui perbandingan lokasi masing-masing klaster.

Pada salah satu bagian di *heatmap.sld*, seperti dapat dilihat kode 5.16, terdapat pengaturan indeks warna (*color hex*) untuk pewarnaan *heat map*. Pewarnaan *heat map* dibedakan menjadi lima warna dengan masing-masing ketebalan warna. Untuk itu, klaster positif menggunakan warna dominan hijau (#008B8B dan #7DFF40), sedangkan klaster negatif menggunakan warna dominan merah (#EE0000 dan #FF7D40).

<i>heatmap.sld</i>	
No	Kode
1	<ColorMap type="ramp" >
2	
3	<ColorMapEntry color="#FFFFFF" quantity="0" label="
4	nodata" opacity="0"/>
5	
6	<ColorMapEntry color="#FFFFFF" quantity="0.02" labe
7	l="nodata" opacity="0"/>
8	
9	<ColorMapEntry color="#EE0000" quantity=".1" label=
10	"nodata"/>
11	
12	<ColorMapEntry color="#FF7D40" quantity=".5" label=
13	"values" />
14	
15	<ColorMapEntry color="#FFFF00" quantity="1.0" label
16	="values" />
17	</ColorMap>

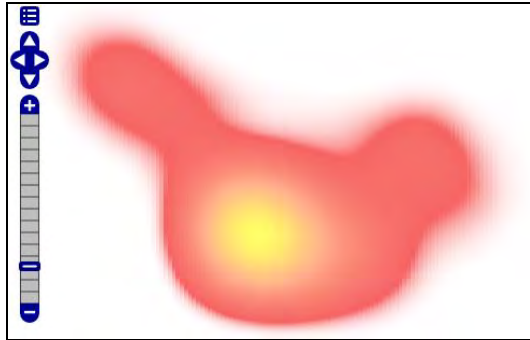
**Kode 5.16. Query Pemberian Warna Heat Map**

Selain untuk pengaturan gradasi warna, *heatmap.sld* juga dapat melakukan pengaturan radius *heatmap* yang diinginkan. Pengaturan awal dari OpenGeo Suite adalah 100. Rentang ini dapat ditingkatkan dengan menambah nilai radius. Semakin besar nilai radius, maka cakupan *heat map* akan semakin luas. Seperti ditunjukkan oleh kode 5.17.

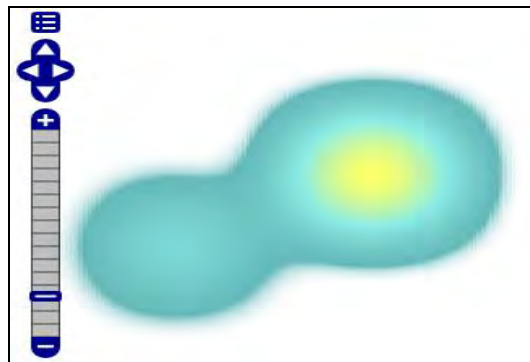
<i>heatmap.sld</i>	
No	Kode
1	<ogc:Function name="parameter">
2	
3	<ogc:Literal>radiusPixels</ogc:Literal>
4	<ogc:Function name="env">
5	<ogc:Literal>radius</ogc:Literal>
6	<ogc:Literal>100</ogc:Literal>
7	</ogc:Function>
8	</ogc:Function>

**Kode 5.17. Query Pengaturan Radius Heat Map**

Untuk menghasilkan peta yang lebih informatif, lapisan peta *heat map* dibedakan menjadi empat jenis yaitu lapisan *heat map* bulan desember klaster positif, lapisan *heat map* bulan desember klaster negatif, lapisan *heat map* bulan april klaster positif, lapisan *heat map* bulan april klaster negatif. Seperti terlihat gambar 5.4 dan gambar 5.5.



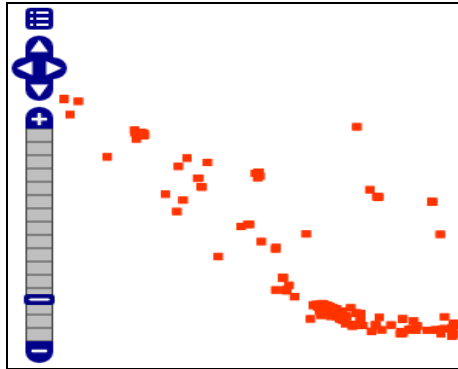
**Gambar 5.4. Klaster Sentimen Negatif**



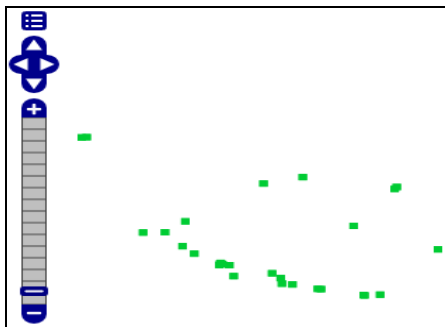
**Gambar 5.5. Klaster Sentimen Positif**

Selain itu juga perlu ada tampilan titik-titik agar memudahkan untuk mengetahui letak pelanggan Telkomsel secara persis. Oleh

karena itu dibuat lapisan peta titik-titik yang dibedakan menjadi empat jenis yaitu lapisan titik-titik bulan desember kluster positif, lapisan titik-titik bulan desember kluster negatif, lapisan titik-titik bulan april kluster positif, lapisan titik-titik bulan april kluster negatif. Seperti terlihat gambar 5.6 dan gambar 5.7.



**Gambar 5.6. Titik-titik Sentimen Negatif**



**Gambar 5.7. Titik-titik Sentimen Positif**

Dengan begitu, lapisan *heat map* untuk kluster positif dan negatif sudah siap ditumpuk dengan peta dasar (peta yang berisikan wilayah-wilayah di seluruh dunia / *base maps*). Untuk mengakses

peta dasar memerlukan koneksi internet. Ada dua peta dasar yang digunakan yaitu peta OpenStreetMap dan peta Bing Aerial With Labels. Seperti ditunjukkan oleh gambar 5.8 dan gambar 5.9.



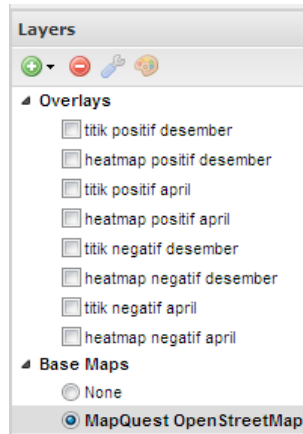
**Gambar 5.8. Peta OpenStreetMap**



**Gambar 5.9. Peta Bing Aerial with Labels**

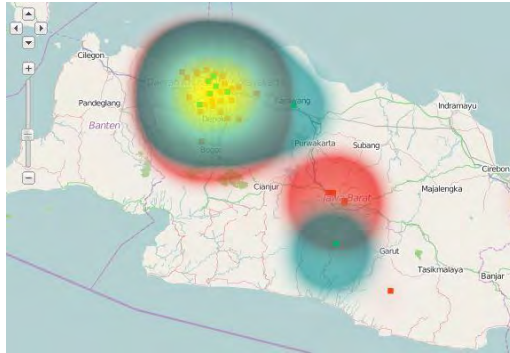
#### 5.2.4. Penyusunan Lapisan di GeoExplorer

Salah satu komponen OpenGeo Suite adalah GeoExplorer. GeoExplorer dapat menggabungkan lapisan *heat map*, lapisan titik-titik dengan peta dasar menjadi sebuah kesatuan peta. Aplikasi ini dapat mengatur susunan lapisan peta yang akan ditampilkan. Baik penempatan lapisan *heat map*, lapisan titik-titik, maupun peta dasar. Gambar 5.10 menunjukkan lapisan-lapisan yang dapat ditampilkan.



**Gambar 5.10. Lapisan Pemetaan**

Di dalam GeoExplorer terdapat menu Overlays dan Base Maps. Overlays berguna untuk pengaturan lapisan apapun berbasis koordinat, sedangkan Base Maps digunakan sebagai lapisan dasar atau peta yang hendak ditempelkan oleh lapisan-lapisan di Overlays. Dengan Overlays memungkinkan memindah tumpukan lapisan ke atas atau ke bawah baik ditampilkan maupun tidak ditampilkan sesuai kebutuhan. Berikut adalah contoh pemetaan data Twitter dalam bentuk *heat map* pada peta OpenStreetMap dan Bing Aerial With Labels. Contoh hasil pemetaan dapat dilihat pada gambar 5.11 dan gambar 5.12.



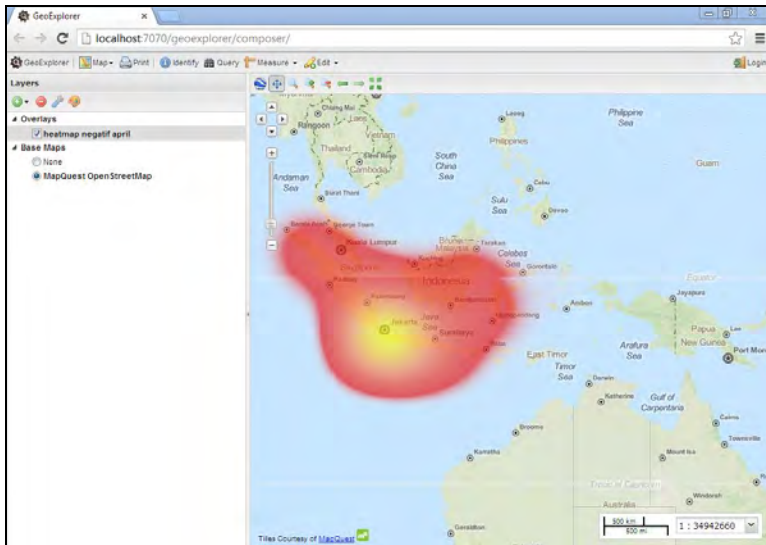
**Gambar 5.11. Tampilan Pemetaan Basis Peta OpenStreetMap**



**Gambar 5.12. Tampilan Pemetaan Basis Peta Bing Aerial with Labels**

### **5.2.5. Pembuatan *Web Page* untuk Pemetaan Data**

Pemetaan data melalui GeoExplorer menghasilkan tampilan bawaan dari OpenGeo Suite. Untuk itu, agar lebih aplikatif, perlu membuat tampilan versi *web page*. Sehingga hasil pemetaan data dipindahkan ke tampilan versi *web page* seperti ditunjukkan gambar 5.13.

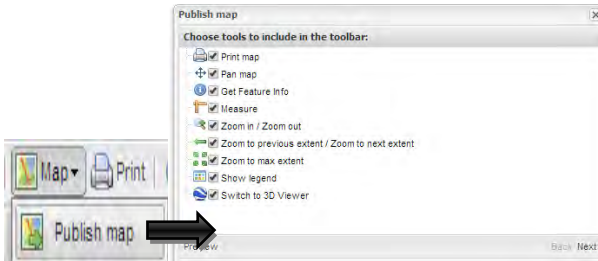


**Gambar 5.13. Tampilan Pemetaan dalam GeoExplorer**

Pembuatan *web page* menggunakan bahasa pemrograman PHP. *Bootstrap Starter* dengan tema *dashboard* digunakan sebagai *open source* dari *Content Management System* (CMS) karena CMS tersebut relatif bersesuaian dengan konten pemetaan data Twitter.

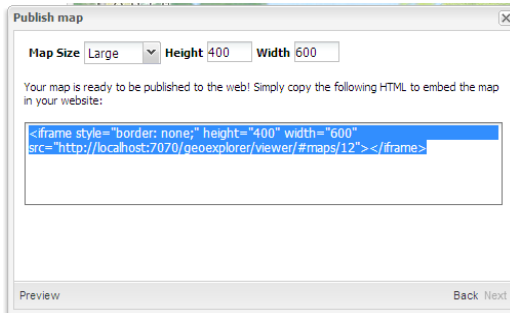
*Web page* membutuhkan *link* dari GeoExplorer yang menghubungkan dengan peta pemetaan data Twitter. Pada dasarnya *web page* dijalankan melalui XAMPP dengan *port* 80, sedangkan GeoExplorer dijalankan melalui OpenGeo Suite dengan *port* 7070. Karena ada perbedaan *port*, maka tidak dapat dihubungkan melalui pengelolaan direktori. Cara lainnya adalah dengan cara memanfaatkan fitur *publish map* dari GeoExplorer. Gambar 5.14 menunjukkan fitur *publish map* sebelum menghasilkan *link* untuk peta.





**Gambar 5.14. Fitur *Publish Map* GeoExplorer**

Fitur *publish map* menghasilkan *html* yang di dalamnya terdapat *port* GeoExplorer. Melalui *publish map* ini juga dapat mengatur tinggi dan lebarnya peta yang akan disebar. *Html* ini dicantumkan ke dalam naskah kode *web page* agar dapat dimunculkan pada *web page*. Gambar 5.15 menunjukkan link peta dari GeoExplorer. Ada beberapa fitur GeoExplorer yang dapat dipilih untuk diaktifkan atau tidak saat ditampilkan di halaman *web page*.



**Gambar 5.15. Link Peta dari GeoExplorer**

*Html* yang diperoleh dari fitur *publish map* disalin ke *class* di salah satu halaman dapat dilihat pada kode 5.18. Sehingga meskipun dijalankan di XAMPP yang mana berbeda *port* dengan OpenGeo Suite, namun peta tetap dapat muncul di halaman *web page*.

No	Kode
1	
2	<iframe style="border: none;" height="400"
3	width="600"src="http://localhost:7070/geoexplorer/vi
4	ewer/#maps/12"></iframe>
5	

**Kode 5.18. Query Link Peta dari GeoExplorer**

*Html* yang diperoleh kemudian dicantumkan pada naskah kode *lihat\_peta.php* di bagian fungsi *class*, maka peta akan muncul di halaman *web page* tersebut. Naskah kode *lihat\_peta.php* dapat dilihat pada kode 5.19.

<i>lihat_peta.php</i>	
No	Kode
1	<html>
2	<body>
3	
4	<div id="wrapper">
5	
6	<nav class="navbar navbar-default navbar-
7	fixed-top" role="navigation" style="margin-bottom:
8	0">
9	<div class="navbar-header">
10	<button type="button"
11	class="navbar-toggle" data-toggle="collapse" data-
12	target=".sidebar-collapse">
13	<span class="sr-only">Toggle
14	navigation</span>
15	<span class="icon-bar"></span>
16	<span class="icon-bar"></span>
17	<span class="icon-bar"></span>
18	</button>
19	<a class="navbar-brand"
20	href="index.html">PT. Telekomunikasi Selular</a>
21	</div>
22	<div class="navbar-default navbar-
23	static-side" role="navigation">
24	<div class="sidebar-collapse">
25	<ul class="nav" id="side-menu">

**Kode 5.19. Query *lihat\_peta.php***

<i>lihat_peta.php</i>	
No	Kode
26	<code>&lt;li class="sidebar-search"&gt;</code>
27	
28	<code>&lt;/li&gt;</code>
29	<code>&lt;li&gt;</code>
30	<code>&lt;a href="index.html"&gt;&lt;i</code>
31	<code>class="fa fa-dashboard fa-fw"&gt;&lt;/i&gt; <b>Profil</b>&lt;/a&gt;</code>
32	<code>&lt;/li&gt;</code>
33	<code>&lt;li&gt;</code>
34	<code>&lt;a href="#"&gt;&lt;i</code>
35	<code>class="fa fa-bar-chart-o fa-fw"&gt;&lt;/i&gt; <b>Tentang</b></code>
36	<b>Peta</b> <span class="fa arrow"></span></a>
37	<code>&lt;ul class="nav nav-</code>
38	<code>second-level"&gt;</code>
39	<code>&lt;li&gt;</code>
40	<code>&lt;a href="lihat_peta.html"&gt;<b>Lihat Peta</b>&lt;/a&gt;</code>
41	<code>&lt;/li&gt;</code>
42	<code>&lt;li&gt;</code>
43	<code>&lt;a href="akurasi.html"&gt;<b>Tingkat Akurasi</b>&lt;/a&gt;</code>
44	<code>&lt;/li&gt;</code>
45	
46	<code>&lt;li&gt;</code>
47	<code>&lt;a href="buat_peta.html"&gt;<b>Alur Pemetaan</b></code>
48	<b>Data Twitter</b> </a>
49	<code>&lt;/li&gt;</code>
50	<code>&lt;/ul&gt;</code>
51	<code>&lt;/li&gt;</code>
52	<code>&lt;/ul&gt;</code>
53	<code>&lt;!-- /#side-menu --&gt;</code>
54	<code>&lt;/div&gt;</code>
55	<code>&lt;!-- /.sidebar-collapse --&gt;</code>
56	<code>&lt;/div&gt;</code>
57	<code>&lt;!-- /.navbar-static-side --&gt;</code>
58	<code>&lt;/nav&gt;</code>
59	
60	<code>&lt;div id="page-wrapper"&gt;</code>
61	<code>&lt;div class="row"&gt;</code>
62	<code>&lt;div class="col-lg-12"&gt;</code>
63	<code>&lt;h1 class="page-header"&gt;</code>
64	<b>Pemetaan Data Twitter</b> </h1>
65	<code>&lt;/div&gt;</code>
66	<code>&lt;!-- /.col-lg-12 --&gt;</code>
67	<code>&lt;/div&gt;</code>
68	<code>&lt;!-- /.row --&gt;</code>
69	<code>&lt;/div&gt;</code>

**Kode 5.18. Query *lihat\_peta.php* (lanjutan)**

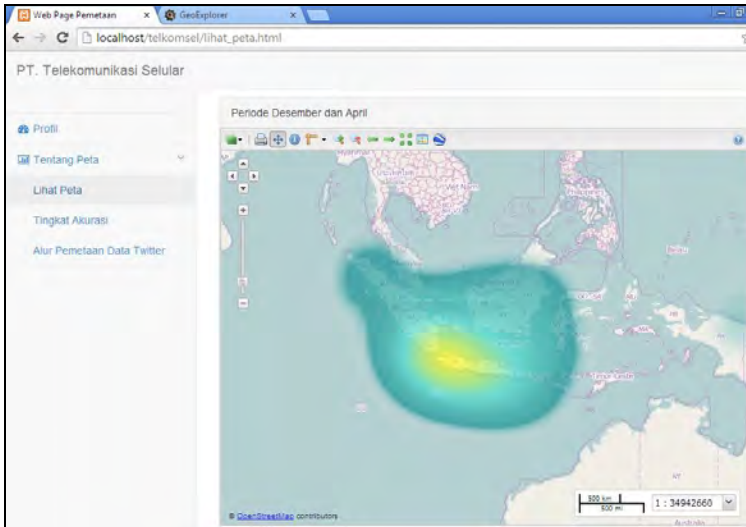
<i>lihat_peta.php</i>	
No	Kode
70	
71	<code>&lt;div class="row"&gt;</code>
72	<code>&lt;div class="col-lg-6"&gt;</code>
73	<code>&lt;p&gt;Pelanggan</code>
74	Telkomsel dapat menyampaikan opininya melalui
75	Twitter. Data tersebut kemudian diambil untuk
76	dipetakan. Berikut adalah hasil pemetaan data
77	Twitter pelanggan Telkomsel: <code>&lt;/p&gt;</code>
78	<code>&lt;div class="panel panel-</code>
79	<code>default"&gt;</code>
80	<code>&lt;div class="panel-heading"&gt;</code>
81	Periode Desember dan
82	April
83	<code>&lt;/div&gt;</code>
84	<code>&lt;!-- /.panel-heading --&gt;</code>
85	
86	<code>&lt;iframe style="border: none;" height="400"</code>
87	<code>width="600"</code>
88	<code>src="http://localhost:7070/geoexplorer/viewer/#maps</code>
89	<code>/12"&gt;&lt;/iframe&gt;</code>
90	
91	<code>&lt;!-- /.panel-body --&gt;</code>
92	<code>&lt;/div&gt;</code>
93	<code>&lt;!-- /.panel --&gt;</code>
94	<code>&lt;/div&gt;</code>
95	<code>&lt;/div&gt;</code>
96	<code>&lt;/div&gt;</code>
97	<code>&lt;/div&gt;</code>
98	
99	<code>&lt;!-- Core Scripts - Include with every page --&gt;</code>
100	<code>&lt;script src="js/jquery-1.10.2.js"&gt;&lt;/script&gt;</code>
101	<code>&lt;script src="js/bootstrap.min.js"&gt;&lt;/script&gt;</code>
102	<code>&lt;script</code>
103	<code>src="js/plugins/metisMenu/jquery.metisMenu.js"&gt;&lt;/sc</code>
104	<code>ript&gt;</code>
105	<code>&lt;!-- SB Admin Scripts - Include with every page</code>
106	<code>--&gt;</code>
107	<code>&lt;script src="js/sb-admin.js"&gt;&lt;/script&gt;</code>

**Kode 5.18. Query *lihat\_peta.php* (lanjutan)**

<i>lihat_peta.php</i>	
No	Kode
108	<code>&lt;!-- Page-Level Demo Scripts - Dashboard - Use</code>
109	<code>for reference --&gt;</code>
110	<code>&lt;script src="js/demo/dashboard-</code>
111	<code>demo.js"&gt;&lt;/script&gt;</code>
112	<code></code>
113	<code>&lt;/body&gt;</code>
114	<code>&lt;/html&gt;</code>
115	<code></code>

**Kode 5.18. Query *lihat\_peta.php* (lanjutan)**

Jika fitur-fitur GeoExplorer diaktifkan, maka fitur-fitur tersebut juga dapat dijalankan di halaman yang telah ditempelkan tersebut. Saat menjalankan peta tersebut, aplikasi OpenGeo Suite harus sudah dalam keadaan sedang berjalan sehingga peta pada *web page* dapat ditampilkan. *Web page* pemetaan yang menampilkan peta GeoExplorer dapat dilihat pada gambar 5.16.



**Gambar 5.16. Tampilan Pemetaan dalam *Web Page***

## **BAB VI**

### **UJI COBA DAN ANALISIS HASIL**

Bab ini menjelaskan mengenai proses uji coba terhadap kinerja aplikasi pengklasifikasi teks *Naïve Bayes* dan uji coba terhadap pemetaan dalam bentuk *heat map*. Pembahasan bab uji coba dan analisis hasil meliputi data uji coba, lingkungan uji coba, skenario uji coba, pelaksanaan dan hasil, serta analisis hasil.

#### **6.1. Data Uji Coba**

Data yang digunakan untuk pelatihan maupun pengujian adalah data yang terkumpul dari proses pengumpulan data menggunakan Twitter API *Stream* pada periode tertentu. Untuk pelatihan, data uji coba meliputi dua periode yang berbeda yaitu periode desember dan periode april. Periode desember dihitung pada tanggal 22-25 Desember 2013. Untuk periode april, data diperoleh pada tanggal 3-6, 10, 12, dan 16-24 April 2014. Sedangkan data pengujian diperoleh pada periode juni yang meliputi tanggal 5-11, 15, 17-18, dan 27-28 Juni 2014.

Data pelatihan dan pengujian masing-masing melalui tahap-tahap praproses data yang sama. Data mentah untuk pelatihan sebelum melalui praproses data ada sebanyak 133.466 data. Kemudian setelah melalui tahap-tahap praproses data, data tersebut menjadi 623 data. Sedangkan data pengujian dari data mentah sebanyak 22.733 menjadi 150 data setelah melalui praproses data. Rincian dapat dilihat pada tabel 6.1.

Dari data pelatihan sebanyak 623, data diproses pemilihan fitur sebagaimana metode *Naïve Bayes*. Fitur yang diperoleh dari data pelatihan tersebut sebanyak 336 fitur. Setiap data dokumen dari data pelatihan mengandung setidaknya sebuah fitur dari total fitur yang diperoleh tersebut. Total fitur ini digunakan untuk melakukan pengujian pada data baru yakni data periode Juni.

**Tabel 6.1. Rincian Jumlah Data Latih dan Data Uji Coba**

Tahap	Data Latih			Data Uji Coba
	Periode Desember	Periode April	Total	Periode Juni
Data awal	8.428	125.038	133.466	22.733
Praproses data:				
1. Pemilihan atribut	8.428	125.038	133.466	22.733
2. Eliminasi data tanpa koordinat	570	2.463	3.033	763
3. Pemilihan data opini	217	561	778	243
4. Pemilihan opini operasional CRM	161	462	623	150

Sedangkan untuk melakukan pengujian pada pemetaan dalam bentuk *heat map*, data yang dibutuhkan adalah data jarak antar titik dari wilayah tertentu. Wilayah yang menjadi pengujian adalah Depok, Jawa Barat. Ada lima titik sebagai *centroid* yaitu A, B, C, D, dan E. Masing-masing besaran nilai jarak ditunjukkan oleh tabel 6.2. Ini digunakan untuk menghitung *hierarchical clustering*. Sehingga perhitungan ini dapat menghasilkan dendogram yang sama dengan hasil *heat map* yang digunakan oleh aplikasi OpenGeo Suite.

**Tabel 6.2. Nilai Jarak untuk Pembuatan Dendogram**

Titik (km)	A	B	C	D	E
A	0	4.7	6.8	7.35	11.7
B	4.7	0	3.9	4.3	7
C	6.8	3.9	0	0.5	7.1
D	7.35	4.3	0.5	0	6.9
E	11.7	7	7.1	6.9	0

## 6.2. Lingkungan Uji Coba

Lingkungan uji coba menjelaskan spesifikasi perangkat keras dan fungsi perangkat lunak yang digunakan dalam tahapan implementasi aplikasi. Spesifikasi perangkat keras lingkungan uji coba yang digunakan dapat dilihat pada tabel 6.3.

**Tabel 6.3. Perangkat Keras Lingkungan Uji Coba**

<b>Perangkat Keras</b>	<b>Spesifikasi</b>
Jenis	Notebook
Processor	Intel® Core™ i7-2630QM CPU @ 2.00GHz (8 CPUs), ~2.0GHz
RAM	4096MB
Hard Disk Drive	68.3GB

Selain itu juga terdapat lingkungan perangkat lunak yang digunakan dalam uji coba. Tabel 6.4 berikut adalah daftar perangkat lunak yang digunakan dalam uji coba.

**Tabel 6.4. Perangkat Lunak Lingkungan Uji Coba**

<b>Perangkat Lunak</b>	<b>Fungsi</b>
Windows 7 Ultimate 32-bit	Sistem Operasi
XAMPP Control Panel v3.1.0, PostGIS	Membuat basis data dan kode program
GeoServer, GeoExplorer	Pemetaan data
Google Chrome, Internet Explorer	Menampilkan Peta
Notepad++	Praproses data
Python 2.7	Pengumpulan data



### 6.3. Skenario Uji Coba

Skenario dilakukan dengan melakukan uji validasi terhadap aplikasi pengklasifikasi teks dan pemetaan dalam bentuk *heat map*. Pengujian dilakukan dengan mengecek secara manual kebenaran dari data yang dihasilkan.

#### 6.3.1. Uji Validasi Aplikasi Pengklasifikasi Teks

Uji validasi untuk aplikasi pengklasifikasi teks terdiri dari tiga skenario yaitu uji akurasi, uji presisi, dan uji *recall*. Sebagaimana dijelaskan pada bagian data uji coba, data pelatihan sebanyak 623 data akan diujikan pada data pengujian sebanyak 150 data dengan menggunakan 133 fitur dari data pelatihan. Berikut masing-masing skenario pengujian:

1. Uji Akurasi

Uji akurasi adalah uji coba pertama yang dilakukan untuk mengetahui keberhasilan ketepatan proses klasifikasi teks. Berdasarkan tabel kontingensi, uji akurasi ditunjukkan oleh persamaan 3 pada bab 2 tinjauan pustaka.

2. Uji Presisi

Uji presisi atau *confidence testing* menunjukkan proporsi dari *Predicted Positive* (PP). Berdasarkan tabel kontingensi, uji presisi dilakukan seperti persamaan 4 pada bab 2 tinjauan pustaka.

3. Uji *Recall*

Uji *recall* atau *sensitivity testing* menunjukkan proporsi dari *Real Positive* (RP). Berdasarkan tabel kontingensi, uji *recall* dilakukan seperti persamaan 5 pada bab 2 tinjauan pustaka.

#### 6.3.2. Uji Validasi Pemetaan dalam Bentuk *Heat Map*

Uji validasi untuk pemetaan dalam bentuk *heat map* dilakukan dengan sebuah skenario uji coba. Pengujian ini dilakukan untuk mengetahui apakah klaster data pemetaan sesuai dengan konsep *hierarchical clustering*. Uji validasi dengan perhitungan *hierarchical clustering* ini dilakukan dengan cara manual.

Pengujian dilakukan dengan menghitung hasil kluster *heat map* dengan konsep *hierarchical clustering* pada wilayah tertentu. Perhitungan kluster tersebut kemudian dibandingkan dengan perhitungan kluster dalam wilayah yang sama namun pada skala yang berbeda. Ini bertujuan untuk mengetahui seberapa jauh cakupan *heat map* berdasarkan densitas data dan ukuran skala peta. Sehingga konsep *heat map* dapat diketahui apakah *single linkage*, *average linkage*, atau *complete linkage*.

Ada tiga jenis skala yang digunakan yaitu 1.000 m, 2.000 m, dan 5.000 m. Obyek yang ditunjukkan masing-masing sama yaitu wilayah Depok, Jawa Barat, dengan lima titik data yang dilakukan kluster. Data yang diujikan juga sama yaitu data yang diperoleh pada periode April dengan klasifikasi sentimen positif.

#### 6.4. Pelaksanaan dan Hasil

Setelah melakukan uji validasi aplikasi pengklasifikasi teks dan pemetaan dalam bentuk *heat map* sesuai dengan skenario uji coba, berikut diperoleh hasil masing-masing pengujian.

##### 6.4.1. Hasil Uji Coba Aplikasi Pengklasifikasi Teks

Pengujian dilakukan dengan menggunakan tabel kontingensi seperti terlihat tabel 6.5. Nilai dari TP, FP, FN, dan TN digunakan dalam masing-masing rumus dari uji akurasi, uji presisi, dan uji *recall*.

Tabel 6.5. Tabel Kontingensi Hasil Klasifikasi Teks

		Manual (kenyataan)	
		Opini Positif	Opini Negatif
Mesin (prediksi)	Opini Positif	TP (20)	FP (38)
	Opini Negatif	FN (2)	TN (90)

## 1. Hasil Uji Akurasi

Nilai uji akurasi berdasarkan hasil uji coba pada tabel 6.4 adalah 73%. Berikut adalah perhitungan uji akurasi dari aplikasi pengklasifikasi teks *Naïve Bayes*.

$$\begin{aligned} \text{Uji Akurasi} &= \frac{TP + TN}{TP + FP + FN + TN} \\ &= \frac{20 + 90}{20 + 38 + 2 + 90} \\ &= 73\% \end{aligned}$$

## 2. Hasil Uji Presisi

Nilai uji presisi berdasarkan hasil uji coba pada tabel 6.4 adalah 34%. Berikut adalah perhitungan uji presisi dari aplikasi pengklasifikasi teks *Naïve Bayes*.

$$\begin{aligned} \text{Uji Presisi} &= \frac{TP}{TP + FP} \\ &= \frac{20}{20 + 38} \\ &= 34\% \end{aligned}$$

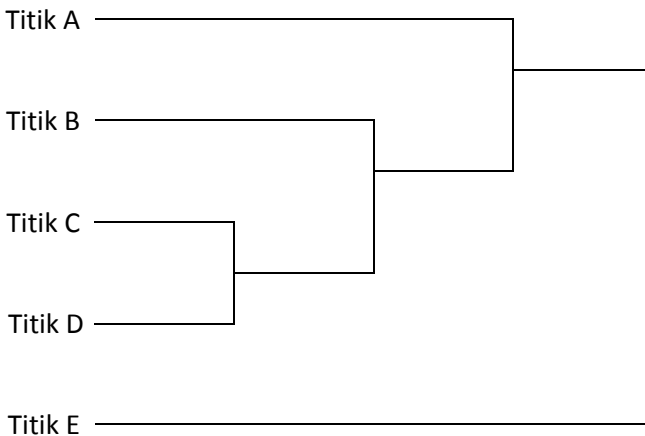
## 3. Hasil Uji Recall

Nilai uji akurasi berdasarkan hasil uji coba pada tabel 6.4 adalah 91%. Berikut adalah perhitungan uji *recall* dari aplikasi pengklasifikasi teks *Naïve Bayes*.

$$\begin{aligned} \text{Uji Recall} &= \frac{TP}{TP + FN} \\ &= \frac{20}{20 + 2} \\ &= 91\% \end{aligned}$$

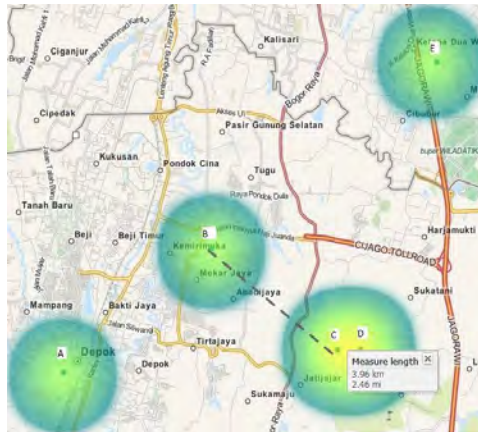
#### 6.4.2. Hasil Uji Coba Pemetaan dalam Bentuk *Heat Map*

Hasil uji coba pemetaan dalam bentuk *heat map* dapat dilakukan dengan menggunakan konsep *hierarchical clustering* dengan *single linkage*. Terdapat lima titik, yaitu A, B, C, D, dan E, yang masing-masing jarak antar titik sudah diketahui dengan menggunakan fitur *measured* pada aplikasi OpenGeo Suite. Jarak diukur dalam satuan kilometer. Kemudian jarak dihitung sebagaimana konsep *single linkage* hingga dihasilkan dendrogram. Hasil dendrogram dari pembuatan klaster *heat map* dari aplikasi OpenGeo Suite adalah sebagai berikut terlihat gambar 6.1.



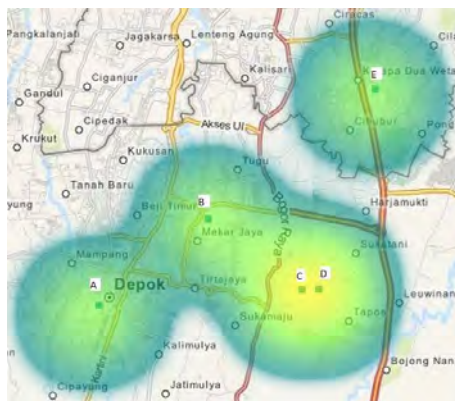
**Gambar 6.1. Dendrogram *Heat Map* yang dihasilkan OpenGeo Suite**

Hasil dendrogram dengan menggunakan *single linkage* tersebut sama dengan yang ditunjukkan oleh gambar 6.2. Gambar tersebut adalah visualisasi pemetaan data dengan menggunakan aplikasi OpenGeo Suite. Terlihat titik C dan titik D adalah klaster yang pertama. Untuk ini, skala yang digunakan adalah 1.000 m.



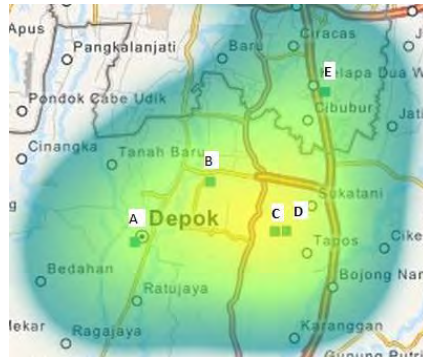
**Gambar 6.2. Klaster *Heat Map* dalam Skala 1.000 m**

Kemudian berdasarkan dendrogram yang sama, klaster berikutnya adalah antara titik CD dan titik B. Jika dilanjutkan klaster menjadi titik BCD dan titik A. Sehingga titik E belum bergabung dengan titik lainnya. Seperti ditunjukkan oleh gambar 6.3. Gambar tersebut merupakan *zoom-out* dari gambar 6.2 dengan skala 2.000 m



**Gambar 6.3. Klaster *Heat Map* dalam Skala 2.000 m**

Berdasarkan dendogram pada gambar 6.1, titik E adalah titik yang terakhir bergabung dengan titik lainnya. Ini sama dengan yang ditunjukkan pada gambar 6.4. Gambar tersebut menunjukkan perluasan skala menjadi 5.000 m



Gambar 6.4. Klaster *Heat Map* dalam Skala 5.000 m

## 6.5. Analisis Hasil

Dari uji coba validasi aplikasi pengklasifikasi teks dan uji coba validasi pemetaan dalam bentuk *heat map*, hasilnya dilakukan analisis terkait faktor-faktor yang mempengaruhi hasil tersebut.

### 6.5.1. Analisis Aplikasi Pengklasifikasi Teks

Peran aplikasi pengklasifikasi teks ini sangat penting sebelum melakukan pemetaan data. Karena data yang dipetakan berdasarkan hasil klasifikasi teks dari aplikasi tersebut. Kebenaran data yang ditampilkan oleh pemetaan akan semakin dapat dipercaya apabila aplikasi pengklasifikasi teks memiliki kemampuan akurasi yang baik. Berdasarkan hasil klasifikasi teks terhadap data uji sebanyak 150 data pada periode Juni, data yang masuk ke dalam klasifikasi sentimen positif sebanyak 58 dari yang seharusnya 22 data, sedangkan di klasifikasi sentimen negatif ada 92 data dari yang seharusnya 128 data. Ini artinya kemampuan aplikasi dalam menentukan klasifikasi masih kurang sempurna. Sehingga hal ini

juga salah satunya yang menyebabkan nilai presisi yang sangat rendah. Dengan melihat kasus-kasus yang terjadi dari hasil klasifikasi teks pada data pengujian, ada dua penyebab mayor terjadinya kesalahan dalam melakukan klasifikasi yaitu penggunaan bahasa tidak baku di Twitter yang sangat beragam dan fenomena terjadinya sarkasme. Berikut adalah pembahasan mengenai kedua hal tersebut. Selain itu juga ada pembahasan analisis terkait nilai uji *recall* yang tinggi sedangkan nilai uji presisinya rendah.

### 6.5.1.1. Keberagaman Kata Tidak Baku di Twitter

Pengklasifikasi teks dengan metode *Naïve Bayes* erat kaitannya dengan pemilihan fitur dari data pelatihan. Tingginya akurasi metode ini juga bergantung dari ketepatan dalam memilih fitur yang mewakili dari data pelatihan tersebut. Sebagaimana dijelaskan sebelumnya, fitur yang didapatkan dari 623 data pelatihan adalah sebanyak 336 fitur. Sebagai berikut contoh dokumen pengujian yang tidak ada fitur yang terakomodasi dari 336 fitur yang tersedia. Oleh karena tidak ada fitur yang tersedia dalam basis data, aplikasi pengklasifikasi teks secara otomatis menggolongkan dokumen tersebut ke klasifikasi positif. Tabel 6.6 menunjukkan beberapa data pengujian yang fiturnya tidak tersedia di basis data.

**Tabel 6.6. Contoh Tweet dengan Bahasa Tidak Baku**

No	Teks	Mesin	Aktual
1	<i>Nomer Telkomselku banyak Sil... Dan bodohnya lagi mau aja aku dikerjain telkompret "@Telkomsel: wirawandanar Masih dengan nomor Sisil"</i>	Positif	Negatif
2	<i>Ini cuma aku aja apa yang semua pake telkomsel hari ini sinyalnya mati idup?</i>	Positif	Negatif

**Tabel 6.6. Contoh Tweet dengan Bahasa Tidak Baku (lanjutan)**

No	Teks	Mesin	Aktual
3	<i>Sinyal H telkomsel byar pet, mosok hrs ke alun2 dulu biar dpt browsing #magelang</i>	Positif	Negatif
4	<i>Dapuk bgt sinyal telkomsel. Aaaaarrrrrrrr &gt;.&lt;</i>	Positif	Negatif
5	<i>Hi Telkomsel, your signal is suck!</i>	Positif	Negatif
6	<i>Malam2 kirim sms lewat telkomsel, sukanya Not Connections Network. Kimalaah!!!</i>	Positif	Negatif
7	<i>Simpati telkomsel is full of sh*t</i>	Positif	Negatif
8	<i>Aihhh telkomsel goblok bgtt!</i>	Positif	Negatif

### 6.5.1.2. Fenomena Sarkasme

Selain karena keterbatasan fitur yang diperoleh dari data pelatihan, persoalan lainnya adalah adanya kalimat sarkasme. Kalimat sarkasme atau kalimat sindiran adalah kalimat yang memiliki kata-kata seolah baik namun memiliki maksud buruk terhadap sesuatu. Seperti terlihat tabel 6.7, beberapa data pengujian mengandung unsur sarkasme yang mana memiliki kata-kata bermakna baik sehingga oleh aplikasi pengklasifikasi teks digolongkan ke klasifikasi sentimen positif.

**Tabel 6.7. Contoh Tweet Fenomena Sarkasme**

No	Teks	Mesin	Aktual
1	<i>@Telkomsel indah dari hongkong</i>	Positif	Negatif
2	<i>Telkomsel sinyalna ampuuuuuun bikin naik darah, untung naik darah ga naik gunung.</i>	Positif	Negatif



**Tabel 6.7. Contoh Tweet Fenomena Sarkasme (lanjutan)**

No	Teks	Mesin	Aktual
3	<i>Aduh pacarku telkomsel lebih perhatian drpd indosat:3 "@IrinAnindiaa: Aduh cici, kamu lbh so sweet drpda pcr indosat aku deh :3</i>	Positif	Negatif
4	<i>RT @yanttimulyantti: Iya de, pacar kedua hahah :-D "@Dahlia26__: Pacar paling setia itu ka telkomsel wkwk</i>	Positif	Negatif

### 6.5.1.3. Analisis Nilai Uji Akurasi

Dari uji coba validasi aplikasi pengklasifikasi teks, hasil uji akurasi memiliki nilai sebesar 73%. Nilai ini berarti apabila ada 10 data yang diujikan, maka kemungkinan 7 data tersebut adalah benar dan sisanya salah. Dalam pengerjaan tugas akhir ini, data yang diujikan ada sebanyak 150. Berdasarkan persentase tersebut, maka kemungkinan kesalahan klasifikasi ada sekitar 40 data uji, sedangkan data yang telah benar diklasifikasikan ada sekitar 110 data uji.

Pengujian akurasi memperhatikan seluruh klasifikasi baik sentimen positif maupun sentimen negatif. Apabila nilai pengujian presisi dan *recall* semakin baik, maka nilai akurasi juga semakin baik. Namun dalam pengujian presisi, nilai yang dihasilkan hanya 34%. Artinya ada banyak data yang seharusnya termasuk dalam klasifikasi sentimen positif, tapi aplikasi memasukkan ke sentimen negatif. Berikut adalah pembahasan lebih lanjut mengenai hasil pengujian presisi dan *recall*.

### 6.5.1.4. Analisis Nilai Uji Presisi yang Rendah dan Nilai Uji Recall yang Tinggi

Hasil untuk uji *recall* dan uji presisi terdapat ketimpangan yang cukup signifikan. Nilai uji *recall* yaitu 91% dan nilai uji presisi yaitu 34% dapat mengindikasikan bahwa aplikasi pengklasifikasi

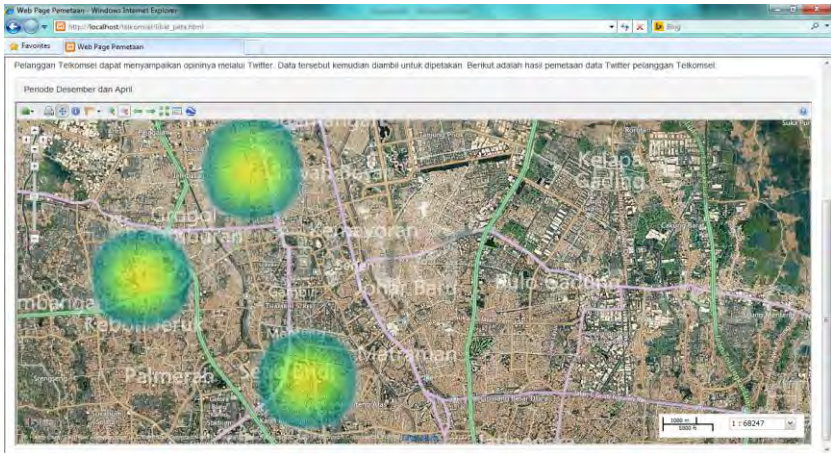
teks lebih baik saat mengklasifikasi data berupa opini negatif daripada opini positif. Sebagaimana pembahasan pada bagian 6.5.1.1 keberagaman bahasa tidak baku di Twitter dan bagian 6.5.1.2 fenomena sarkasme, data yang dicontohkan seluruhnya termasuk *Predicted Positive* (PP). Hal ini mempengaruhi rendahnya nilai uji presisi. Oleh karena itu, jika kedua persoalan tersebut dapat diatasi, maka nilai uji presisi juga akan meningkat.

Tingginya nilai uji *recall* mengindikasikan bahwa aplikasi pengklasifikasi teks relatif baik dalam mengklasifikasikan data pengujian berupa opini negatif. Hal ini dikarenakan kata kunci negatif lebih banyak daripada kata kunci positif sehingga suatu dokumen lebih rentan diklasifikasikan ke kelas negatif. Oleh karena ketepatan klasifikasi untuk opini negatif yang tinggi, maka manfaatnya petugas tidak harus melihat keseluruhan data opini negatif sekadar untuk memastikan bahwa tidak ada opini negatif yang tidak tertangani.

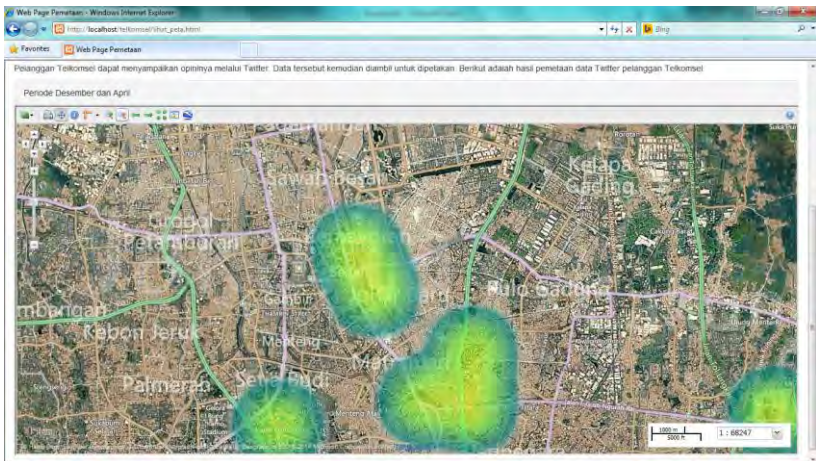
### **6.5.2. Analisis Pemetaan dalam Bentuk *Heat Map***

Ada tiga konsep *hierarchical clustering* yang dapat digunakan untuk membuat kluster yaitu *single linkage*, *average linkage*, dan *complete linkage*. Berdasarkan hasil uji coba validasi pemetaan data dalam bentuk *heat map*, kluster *heat map* yang dihasilkan oleh aplikasi OpenGeo Suite merupakan *hierarchical clustering* dengan kemiripan kluster *single linkage*. *Single linkage* adalah pembuatan kluster dengan berdasarkan dua titik kemiripan terdekat dengan kluster yang berbeda.

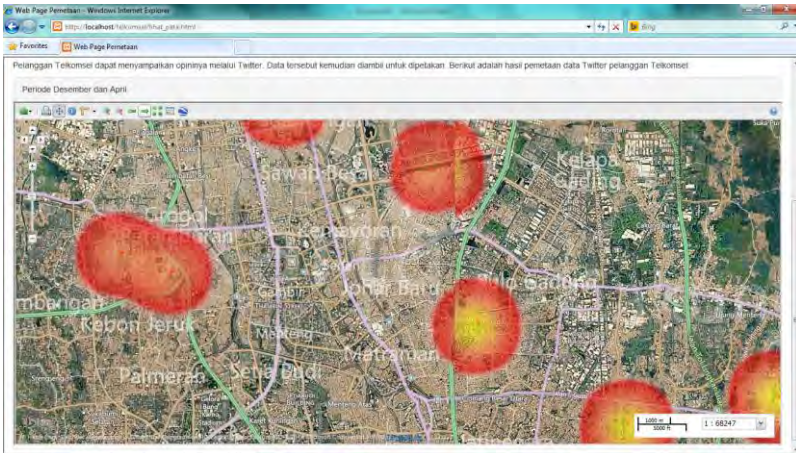
*Single linkage* atau pemetaan dalam bentuk *heat map* yang dihasilkan pada tugas akhir ini memiliki kelebihan dan kekurangan dalam penggunaannya. Kelebihannya adalah dapat mengatasi bentuk yang tidak bulat panjang (*non-elliptical shapes*). Sedangkan kekurangannya adalah sensitif terhadap adanya riuh (*noise*) dan *outlier*. Berikut gambar 6.5, gambar 6.6, gambar 6.7, dan gambar 6.8 adalah hasil pemetaan data berbasis *web page* di wilayah Jakarta Pusat pada periode tertentu.



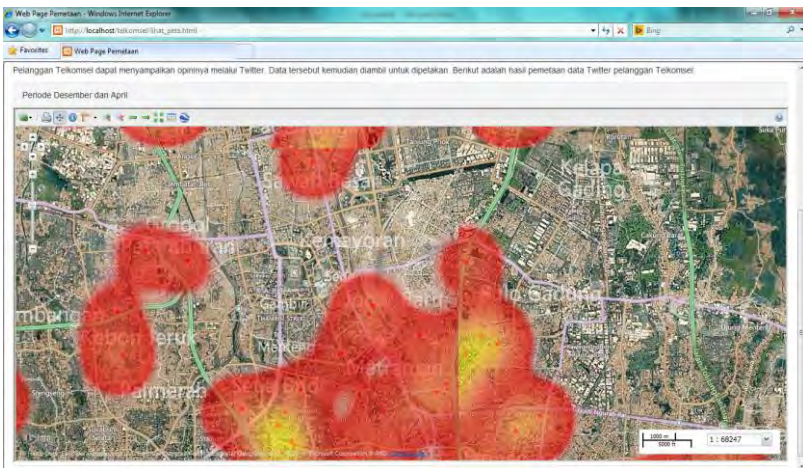
**Gambar 6.5. Pemetaan Data Sentimen Positif di Periode Desember**



**Gambar 6.6. Pemetaan Data Sentimen Positif di Periode April**



**Gambar 6.7. Pemetaan Data Sentimen Negatif di Periode Desember**



**Gambar 6.8. Pemetaan Data Sentimen Negatif di Periode April**

*Halaman ini sengaja dikosongkan*

## **BAB VII KESIMPULAN**

Bab ini berisi kesimpulan dari hasil klasifikasi opini positif dan negatif oleh aplikasi pengklasifikasi teks berbasis *Naive Bayes* dan pemetaan hasil klasifikasi tersebut dalam bentuk *heat map*. Di bawah ini diberikan kesimpulan yang dapat diperoleh setelah menyelesaikan perancangan dan pembangunan aplikasi pengklasifikasi teks dan pemetannya dalam bentuk *heat map*.

1. Aplikasi pengklasifikasi teks berbasis *Naive Bayes* yang dibangun dengan menggunakan 623 data twitter pelanggan PT. Telekomunikasi Seluar dalam periode bulan Desember 2013 dan April 2014 telah diujicobakan dengan menggunakan 150 data twitter pelanggan dalam periode bulan Juni 2014. Hasil uji coba aplikasi memberikan hasil akurasi dan recall yang memadai dengan nilai berturut-turut sebesar 73% dan 91%. Tetapi, hasil presisi klasifikasi memberikan hasil yang kurang memuaskan dengan nilai sebesar 34%. Hasil presisi yang relatif rendah ini kemungkinan besar dikarenakan cukup banyaknya kata yang tidak baku dan ambigu dalam data twitter yang digunakan dalam membangun pengklasifikasi, sehingga sulit untuk diklasifikasi dengan benar.
2. Penggunaan *heat map* untuk memetakan posisi koordinat pelanggan yang opininya telah diklasifikasi akan memudahkan pihak pengguna untuk mengamati tingkat kepadatan opini pelanggan berdasarkan visualisasi gradasi warna peta yang disajikan pada radius tertentu dari satu titik koordinat pengamatan yang diinginkan. Warna hijau kekuningan mengindikasikan kepadatan opini positif yang tinggi, sedang warna merah kekuningan mengindikasikan kepadatan opini negatif yang tinggi. Namun demikian, oleh

karena peta yang dihasilkan oleh *heat map* didasarkan pada metode klasterisasi berbasis aglomeratif, maka pengelompokan titik-titik yang digabungkan dalam hasil pemetaan menjadi sensitif terhadap derau dan pencilan, sehingga struktur pengelompokan data yang terbentuk dapat menjadi kurang tepat.

## DAFTAR PUSTAKA

- BoundlessGeo. (2014). *Tutorial*. Diambil kembali dari BoundlessGeo Web site: <http://workshops.boundlessgeo.com/#tutorials>
- Calvin, & Setiawan, J. (2014). Using Text Mining to Analyze Mobile Phone Provider Service Quality (Case Study: Social Media Twitter). *International Journal of Machine Learning and Computing*, 106-109.
- Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter Brand Sentiment Analysis: A Hybrid System using n-gram Analysis and Dynamic Artificial Neural Network. *ScienceDirect*, 6268.
- Ikonomakis, M., Kotsiantis, S., & Tampakas, V. (2005). Text Classification Using Machine Learning Techniques. *WSEAS Transactions on Computers Issue 8 Volume 4*, 966-974.
- Kontopoulos, E., Berberidis, C., Dergiades, T., & Bassiliades, N. (2013). Ontology-based Sentiment Analysis of Twitter Posts. *ScienceDirect*, 4066.
- Lestari, N. M., Putra, K. G., & Chayawan, K. A. (2013). Personality Types Classification for Indonesian Text in Partners Searching Website Using Naive Bayes Methods. *International Journal of Computer Sciences Issues*, 1-8.
- Liu, B. (2012). Sentiment Analysis: a Fascinating Problem. Dalam B. Liu, *Sentiment Analysis and Opinion Mining* (hal. 7-8). Chicago: Morgan & Claypool Publishers.



- Maharani, W. (2013). Microblogging Sentiment Analysis with Lexical Based and Machine Learning Approaches. *IEEE*, 439-443.
- Matthew C. Wagner, S. V. (2013). Application Programming Interfaces (API) A Primer and Discussion of Oracle America vs Google. Dalam S. V. Matthew C. Wagner, *Application Programming Interfaces (API) A Primer and Discussion of Oracle America vs Google* (hal. 3). America: Copyright Law Committee of the American Intellectual Property Law Association.
- Mishra, A., & Mishra, D. (2009). Customer Relationship Management: Implementation Process Perspective. *Acta Polytechnica Hungarica*, 83-99.
- Operator Seluler. (2014). *News*. Diambil kembali dari Operator Seluler: <http://operatorseluler.com/9604/telkomsel-ukir-sejarah-baru-dengan-100-juta-pelanggan/>
- PT. Telekomunikasi Selular. (2012). *Annual Report: Building the Foundation Toward the Digital Era*. Jakarta.
- Sentdex. (2013, September 23). *How to use the Twitter API 1.1 To Stream Tweets in Python*. Diambil kembali dari Sentdex Web site: <http://sentdex.com/sentiment-analysisbig-data-and-python-tutorials-algorithmic-trading/how-to-use-the-twitter-api-1-1-to-stream-tweets-in-python/>
- Tan, P.-N., Steinbach, M., & Kumar, V. (2006). Introduction to Data Mining. Dalam P.-N. Tan, M. Steinbach, & V. Kumar, *Cluster Analysis: Basic Concepts and Algorithms* (hal. 487-493). Boston: Pearson International Edition.

- Thomson Reuters Life Sciences. (2014). *Heat What is a Heat Map*. Diambil kembali dari Thomson Reuters Life Sciences Web site: [custom-analytics.thomsonreuterslifesciences.com/SpotfireWeb/Help/dxpwebclient/heat\\_what\\_is\\_a\\_heat\\_map.htm](http://custom-analytics.thomsonreuterslifesciences.com/SpotfireWeb/Help/dxpwebclient/heat_what_is_a_heat_map.htm)
- Vienna. (2013, Agustus Selasa). *Heatmap Explanation: Hierarchical Clustering*. Diambil kembali dari R Development Core Team: <http://www.hiv.lanl.gov/content/sequence/HEATMAP/help.html>
- Yahya, Y. (2013). *Pengaruh Customer Relationship Management (CRM) dalam Meningkatkan loyalitas Pelanggan*. Jakarta: Universitas Budi Luhur.

*Halaman ini sengaja dikosongkan*

## RIWAYAT PENULIS



Mochamad Nizar Palefi Ma'ady lahir di Bojonegoro pada tanggal 08 November 1991. Anak ketiga dari lima putra bersaudara ini merupakan anak dari Bapak Mochamad Ali dan Ibu Lilik Malikhah. Penulis menempuh masa pendidikan formal di SDN Kadipaten 1 Bojonegoro, SMPN 6 Bojonegoro, dan SMAN 1 Bojonegoro

Pada tahun 2010, penulis melanjutkan studi S1 di Jurusan Sistem Informasi FTIf - Institut Teknologi Sepuluh November (ITS). Di jurusan tersebut penulis terdaftar dengan NRP 5210100074

Di Jurusan Sistem Informasi ini penulis mengambil bidang minat *Decision Support System and Business Intelligence* untuk tugas akhir. Semasa menempuh pendidikan S1 ini penulis juga menjadi asisten praktikum untuk mata kuliah Sistem Fungsional Bisnis I, Sistem Operasi, dan Pengantar Sistem Informasi. Selain itu, penulis juga mengikuti organisasi mahasiswa jurusan, yaitu Himpunan Mahasiswa Sistem Informasi dan dipercaya menjadi staf kewirausahaan pada tahun kedua, staf pengembangan minat dan bakat, staf *Steering Committee*, dan ketua *Steering Committee* FTIf pada tahun ketiga.

Penulis dapat dihubungi melalui email [mnizarpm@gmail.com](mailto:mnizarpm@gmail.com) untuk keperluan penelitian.

*Halaman ini sengaja dikosongkan*

## LAMPIRAN A - DATA FITUR PELATIHAN

Tabel A-8. Fitur Pilihan Untuk Data Pelatihan

idteks	teks	Pscore	Nscore
1460001	rame	3	4
1460002	asik	1	0
1460003	thanks	5	4
1460004	kuat	4	3
1460005	kurang	3	4
1460006	dikibuli	0	1
1460007	gak	58	59
1460008	gaada	0	1
1460009	kek	4	5
1460010	mulu	27	28
1460011	manteep	1	0
1460012	songong	0	1
1460013	gabisa	3	4
1460014	mantep	1	0
1460015	jauuuuh	0	1
1460016	jelek	18	19
1460017	aman	4	3
1460018	ilang	22	23
1460019	gagal	9	10
1460020	tidak	17	18
1460021	lemah	2	3
1460022	susah	15	16
1460023	lemot	51	52
1460024	cukkk	0	1

**Tabel A-9. Fitur Pilihan Untuk Data Pelatihan (lanjutan)**

idteks	teks	Pscore	Nscore
1460025	lelah	1	2
1460026	nyesek	0	1
1460027	maacih	1	0
1460028	galau	3	4
1460029	tengkyu	1	0
1460030	fakir	0	1
1460031	lelet	15	16
1460032	makasih	7	6
1460033	tdk	4	5
1460034	ganteng	1	0
1460035	emosi	2	3
1460036	cinta	1	0
1460037	lancak	1	0
1460038	aduh	1	2
1460039	sabar	0	1
1460040	belum	2	3
1460041	kesel	6	7
1460042	eror	1	2
1460043	thankuuu	1	0
1460044	nyangkut	0	1
1460045	ngeseli	2	3
1460046	iyalah	1	0
1460047	dasar	0	1
1460048	m*h*l	0	1
1460049	selingkuhannya	0	1
1460050	tinggal	0	1

**Tabel A-10. Fitur Pilihan Untuk Data Pelatihan (lanjutan)**

idteks	teks	Pscore	Nscore
1460051	benci	1	2
1460052	wedus	0	1
1460053	tahayul	1	2
1460054	ngadat	4	5
1460055	huft	2	3
1460056	weird	0	1
1460057	kerjinan	0	1
1460058	lambat	0	1
1460059	ganggu	20	21
1460060	bejek	0	1
1460061	angin	0	1
1460062	antri	4	5
1460063	trust	1	0
1460064	togel	5	6
1460065	boring	0	1
1460066	kirain	2	3
1460067	sediih	0	1
1460068	salah	6	7
1460069	cepat	3	2
1460070	kecewa	2	3
1460071	hati	6	5
1460072	sucks	2	3
1460073	wrong	0	1
1460074	ribet	0	1
1460075	kenceng	3	2
1460076	php	0	1



**Tabel A-11. Fitur Pilihan Untuk Data Pelatihan (lanjutan)**

idteks	teks	Pscore	Nscore
1460077	seneng	1	0
1460078	sos	6	7
1460079	tagihan	0	1
1460080	kebangeten	0	1
1460081	turun	3	4
1460082	asem	0	1
1460083	padahal	1	2
1460084	cacad	0	1
1460085	pending	8	9
1460086	segni	0	1
1460087	jebol	2	3
1460088	keluhan	0	1
1460089	mendingan	0	1
1460090	lenyap	3	4
1460091	fuck	2	3
1460092	susaah	0	1
1460093	kacau	2	3
1460094	lancar	13	12
1460095	migrasi	1	0
1460096	selain	1	2
1460097	awet	1	0
1460098	memusingkan	0	1
1460099	bagus	5	4
1460100	knp	1	2
1460101	mahal	8	9
1460102	mesti	3	4

**Tabel A-12. Fitur Pilihan Untuk Data Pelatihan (lanjutan)**

idteks	teks	Pscore	Nscore
1460103	yahut	1	0
1460104	hina	0	1
1460105	larang	1	2
1460106	ngambek	0	1
1460107	hebat	0	1
1460108	bener	2	3
1460109	spam	0	1
1460110	smsin	2	3
1460111	love	2	1
1460112	ampun	0	1
1460113	hancuur	0	1
1460114	damn	0	1
1460115	sedot	1	2
1460116	potong	5	6
1460117	tak	2	3
1460118	boro	2	1
1460119	paraaah	0	1
1460120	makanya	5	4
1460121	stupid	0	1
1460122	huff	0	1
1460123	beres	1	0
1460124	gapernah	1	2
1460125	garuda	1	0
1460126	murah	1	0
1460127	disconect	0	1
1460128	kalah	0	1

**Tabel A-13. Fitur Pilihan Untuk Data Pelatihan (lanjutan)**

idteks	teks	Pscore	Nscore
1460129	esemes	1	2
1460130	sorry	0	1
1460131	aigoo	0	1
1460132	rencok	0	1
1460133	ribut	0	1
1460134	parah	8	9
1460135	cacat	2	3
1460136	unmood	0	1
1460137	poop	0	1
1460138	nyangka	1	0
1460139	error	15	16
1460140	untungnya	1	0
1460141	terjangkau	1	0
1460142	galigot	0	1
1460143	setia	4	3
1460144	reliable	2	1
1460145	ngbari	0	1
1460146	bermasalah	2	3
1460147	anjing	0	1
1460148	kacrut	0	1
1460149	wuzz	2	1
1460150	taken	0	1
1460151	disappearing	0	1
1460152	sibuk	0	1
1460153	failed	1	2
1460154	baik	1	0

**Tabel A-14. Fitur Pilihan Untuk Data Pelatihan (lanjutan)**

idteks	teks	Pscore	Nscore
1460155	complain	4	5
1460156	shit	0	1
1460157	bosok	1	2
1460158	kimak	0	1
1460159	aneh	0	1
1460160	ditarik	0	1
1460161	risih	0	1
1460162	pergi	0	1
1460163	keren	1	0
1460164	padamu	1	0
1460165	goa	0	1
1460166	tahe	0	1
1460167	ngemceeh	0	1
1460168	nunggu	0	1
1460169	atuh	0	1
1460170	kuaat	1	0
1460171	buruk	1	2
1460172	jancok	0	1
1460173	busuk	4	5
1460174	prhatian	1	2
1460175	nyedot	0	1
1460176	seperhatian	0	1
1460177	kampret	1	2
1460178	bengaak	0	1
1460179	ngarep	0	1
1460180	gannggu	1	2

**Tabel A-15. Fitur Pilihan Untuk Data Pelatihan (lanjutan)**

idteks	teks	Pscore	Nscore
1460181	abis	4	5
1460182	istighfar	0	1
1460183	payah	0	1
1460184	bothering	0	1
1460185	ngertiin	1	0
1460186	rudet	1	2
1460187	tokcer	1	0
1460188	naik	4	3
1460189	ngelek	0	1
1460190	putus	1	2
1460191	bego	0	1
1460192	ngeyel	0	1
1460193	ganti	6	5
1460194	overprotektif	0	1
1460195	tiada	0	1
1460196	bosen	0	1
1460197	took	0	1
1460198	ganguan	0	1
1460199	menghambat	0	1
1460200	resek	0	1
1460201	terjaga	1	0
1460202	berfungsinya	0	1
1460203	kendala	0	1
1460204	ngebut	1	0
1460205	gngguan	3	4
1460206	semplak	0	1

**Tabel A-16. Fitur Pilihan Untuk Data Pelatihan (lanjutan)**

idteks	teks	Pscore	Nscore
1460207	sepinya	0	1
1460208	babi	0	1
1460209	bgus	1	0
1460210	nyesal	0	1
1460211	bangsat	0	1
1460212	terlalu	0	1
1460213	anjiiirrrr	0	1
1460214	trouble	0	1
1460215	rugi	1	2
1460216	terkuras	1	2
1460217	stress	0	1
1460218	kecu	0	1
1460219	miris	0	1
1460220	dibangunin	0	1
1460221	unyu	0	1
1460222	racun	0	1
1460223	mendadak	1	2
1460224	full	2	1
1460225	lama	1	2
1460226	tethering	1	0
1460227	boros	0	1
1460228	menjamin	2	3
1460229	gajelas	0	1
1460230	berjaya	1	0
1460231	heran	1	0
1460232	ngemeng	0	1

**Tabel A-17. Fitur Pilihan Untuk Data Pelatihan (lanjutan)**

idteks	teks	Pscore	Nscore
1460233	phpin	0	1
1460234	promo	1	2
1460235	nyeseg	0	1
1460236	bebulik	1	0
1460237	rajin	0	1
1460238	stop	0	1
1460239	gembel	0	1
1460240	komplain	0	1
1460241	disms	0	1
1460242	gagah	2	1
1460243	terharu	2	3
1460244	ancur	0	1
1460245	bete	2	3
1460246	sharusnya	0	1
1460247	ngeluh	0	1
1460248	takut	0	1
1460249	cabal	0	1
1460250	berkelahi	0	1
1460251	service	3	4
1460252	caper	0	1
1460253	anjlok	0	1
1460254	doang	2	1
1460255	wafat	0	1
1460256	beralih	1	0
1460257	mengecewakan	1	2
1460258	bemsalah	0	1

**Tabel A-18. Fitur Pilihan Untuk Data Pelatihan (lanjutan)**

idteks	teks	Pscore	Nscore
1460259	bapuk	1	2
1460260	menurun	0	1
1460261	bacut	0	1
1460262	bosan	0	1
1460263	kok	32	33
1460264	mahaaaaaaal	0	1
1460265	jomblo	2	3
1460266	suddenly	1	2
1460267	kepake	0	1
1460268	tertarik	0	1
1460269	perhatian	15	16
1460270	pale	2	3
1460271	ngamok	0	1
1460272	kurang	5	6
1460273	gile	0	1
1460274	prasaan	0	1
1460275	awalnya	0	1
1460276	menyiksa	0	1
1460277	tolong	6	7
1460278	pancen	0	1
1460279	tiada	0	1
1460280	bingit	1	2
1460281	efek	2	3
1460282	pantas	0	1
1460283	kayaknya	2	3
1460284	giliran	3	4



**Tabel A-19. Fitur Pilihan Untuk Data Pelatihan (lanjutan)**

idteks	teks	Pscore	Nscore
1460285	kenapa	0	1
1460286	ngeten	1	2
1460287	akhir2	2	3
1460288	tragis	1	2
1460289	melebihi	3	4
1460290	cuy	1	2
1460291	amit	0	1
1460292	mesti	5	6
1460293	kayanya	0	1
1460294	maket	2	3
1460295	coba	9	10
1460296	ndaada	0	1
1460297	kebawah	0	1
1460298	menengah	0	1
1460299	ndak	5	6
1460300	pacar	8	9
1460301	lola	2	3
1460302	diacount	1	2
1460303	smpai	1	2
1460304	drop	0	1
1460305	telkomsex	0	1
1460306	biasanya	5	6
1460307	jadi	17	18
1460308	adanya	2	3
1460309	bkn	2	3
1460310	astaga	1	2

**Tabel A-20. Fitur Pilihan Untuk Data Pelatihan (lanjutan)**

idteks	teks	Pscore	Nscore
1460311	sialan	2	3
1460312	- -	38	39
1460313	=))	2	1
1460314	:/	0	1
1460315	:(	19	20
1460316	:D	2	1
1460317	x_x	2	3
1460318	- _____ -	0	1
1460319	:)	5	4
1460320	:-{	0	1
1460321	:*	2	1
1460322	- _____ -	0	1
1460323	- _ _ -	0	1
1460324	- _ _ -	1	2
1460325	:v	1	0
1460326	- _____ -	0	1
1460327	B-)	1	0
1460328	=	1	2
1460329	:3	1	0
1460330	x_x	3	4
1460331	(y)	1	0
1460332	(n)	0	1
1460333	!!!!	6	7
1460334	LELET	0	1
1460335	NGEYEL	0	1
1460336	({})	2	3

