



SKRIPSI

**PERBANDINGAN METODE WARD'S, DBSCAN,
DAN K-MEANS UNTUK KLASTERISASI ADUAN
PELAYANAN MASYARAKAT DI
DISPENDUKCAPIL KOTA SURABAYA**

MARETA PUTRI WARDHANA
NPM 21083010002

DOSEN PEMBIMBING
Kartika Maulida Hindrayani, S.Kom., M.Kom.
Amri Muhaimin, S.Stat., M.Stat., M.S.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2026**

LEMBAR PENGESAHAN

PERBANDINGAN METODE WARD'S, DBSCAN, DAN K-MEANS UNTUK KLASTERISASI ADUAN PELAYANAN MASYARAKAT DI DISPENDUKCAPIL KOTA SURABAYA

Oleh:
MARETA PUTRI WARDHANA
NPM. 21083010002

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur pada Tanggal 25 Februari 2026:

Menyetujui,

Kartika Maulida Hindrayani, S.Kom., M.Kom.
NIP. 199209092022032009



(Pembimbing I)

Amri Muhaimin, S.Stat., M.Stat., M.S.
NIP. 199507232024061002



(Pembimbing II)

Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
NIP. 198303102021211006



(Ketua Penguji)

Shindi Shella May Wara, S.Si., M.Stat.
NIP. 199605182024062003



(Penguji I)

Mengetahui,

Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT.

NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

PERBANDINGAN METODE WARD'S, DBSCAN, DAN K-MEANS UNTUK KLASTERISASI ADUAN PELAYANAN MASYARAKAT DI DISPENDUKCAPIL KOTA SURABAYA

Oleh:
MARETA PUTRI WARDHANA
NPM. 21083010002

Telah disetujui untuk mengikuti Ujian Skripsi

Menyetujui,

Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer

Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU., Asean Eng.
NIP. 19801205 200501 1 002

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Mareta Putri Wardhana
NPM : 21083010002
Program : Sarjana (S1)
Program Studi : Sains Data
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila di kemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, 4 Maret 2026
Yang Membuat Pernyataan,



MARETA PUTRI WARDHANA
NPM. 21083010002

ABSTRAK

Nama Mahasiswa / NPM : Mareta Putri Wardhana / 21083010002
Judul Skripsi : Perbandingan Metode Ward's, DBSCAN, Dan K-Means Untuk Klasterisasi Aduan Pelayanan Masyarakat Di Dispendukcapil Kota Surabaya
Dosen Pembimbing : 1. Kartika Maulida Hindrayani, S.Kom., M.Kom.
2. Amri Muhaimin, S.Stat., M.Stat., M.S.

Pengelompokan data teks merupakan salah satu teknik dalam *Text Mining* yang digunakan untuk mengidentifikasi pola pada data tidak terstruktur. Dalam pelayanan publik, aduan masyarakat umumnya disampaikan dalam bentuk teks deskripsi keluhan yang beragam sehingga sulit dianalisis secara manual. Kondisi ini juga terjadi pada data aduan di Dispendukcapil Kota Surabaya, di mana variasi isi dan struktur bahasa menyebabkan proses identifikasi pola permasalahan menjadi kurang efektif. Oleh karena itu, diperlukan metode klasterisasi untuk mengelompokkan aduan berdasarkan kemiripan isi secara otomatis. Penelitian ini bertujuan membandingkan kinerja *Ward's Method*, DBSCAN, dan K-Means dalam melakukan klasterisasi aduan masyarakat serta menentukan metode yang paling optimal berdasarkan validitas struktur klaster yang dihasilkan. Tahapan penelitian meliputi *preprocessing* teks berupa *cleaning*, *tokenizing*, *normalization*, *stopword removal*, dan *stemming*, kemudian dilakukan pembobotan menggunakan TF-IDF untuk mengubah teks menjadi representasi numerik. Proses klasterisasi diterapkan menggunakan *Ward's Method* berbasis hierarchical agglomerative dengan jarak *Euclidean*, DBSCAN berbasis kepadatan dengan penanganan *noise*, serta K-Means berbasis partisi dengan penentuan jumlah klaster optimal menggunakan *Variance Ratio Criterion* (VRC). Kualitas klaster yang terbentuk dievaluasi menggunakan *Silhouette Coefficient* dan *Index Davies-Bouldin* untuk mengukur tingkat kekompakan dan pemisahan antar klaster. Hasil penelitian menunjukkan bahwa Ward menghasilkan 2 klaster, DBSCAN menghasilkan 10 klaster dan 1 klaster noise, dan K-Means menghasilkan 3 klaster. Berdasarkan evaluasi, Ward memperoleh nilai *Silhouette Coefficient* tertinggi sebesar 0,76 dan *Index Davies-Bouldin* terendah sebesar 1,02, dibandingkan DBSCAN yang menghasilkan nilai *Silhouette* 0,52, dan IDB 1,12, serta K-Means yang menghasilkan nilai *Silhouette* 0,49, dan IDB 3,47. Temuan ini menunjukkan bahwa *Ward's Method* merupakan metode yang paling sesuai dengan karakteristik data aduan pelayanan masyarakat dan dapat digunakan sebagai pendekatan analisis untuk mendukung peningkatan kualitas pelayanan publik.

Kata Kunci: *Text Mining, Klasterisasi, Ward's Method, DBSCAN, K-means*

ABSTRACT

Student Name / NPM : Mareta Putri Wardhana / 21083010002
Undergraduate thesis title : *Comparison Of Ward's Method, Dbscan, And K-Means For Clustering Public Service Complaints At Dispendukcapil Surabaya City*
Advisors : 1. Kartika Maulida Hindrayani, S.Kom., M.Kom.
2. Amri Muhaimin, S.Stat., M.Stat., M.S.

Text data clustering is one of the techniques in Text Mining used to identify patterns in unstructured data. In public services, community complaints are generally conveyed in the form of diverse textual descriptions of complaints, making them difficult to analyze manually. This condition also occurs in complaint data at the Department of Population and Civil Registration (Dispendukcapil) of Surabaya City, where variations in content and language structure cause the process of identifying problem patterns to be less effective. Therefore, clustering methods are needed to group complaints based on content similarity automatically. This study aims to compare the performance of Ward's Method, DBSCAN, and K-Means in clustering public complaints and to determine the most optimal method based on the validity of the resulting cluster structure. The research stages include text preprocessing consisting of cleaning, tokenizing, normalization, stopword removal, and stemming, followed by TF-IDF weighting to transform text into numerical representation. The clustering process is implemented using Ward's Method based on hierarchical agglomerative clustering with Euclidean distance, DBSCAN based on density with noise handling, and K-Means based on partitioning with the optimal number of clusters determined using the Variance Ratio Criterion (VRC). The quality of the resulting clusters is evaluated using the Silhouette Coefficient and the Davies-Bouldin Index to measure the level of compactness and separation between clusters. The results show that Ward produces 2 clusters, DBSCAN produces 10 clusters and 1 cluster as noise, and K-Means produces 3 clusters. Based on the evaluation, Ward obtains the highest Silhouette Coefficient value of 0.76 and the lowest Davies-Bouldin Index value of 1.02, compared to DBSCAN which produces a Silhouette value of 0.52 and DBI of 1.12, and K-Means which produces a Silhouette value of 0.49 and DBI of 3.47. These findings indicate that Ward's Method is the most suitable method for the characteristics of public service complaint data and can be used as an analytical approach to support improvements in public service quality.

Keywords: *Text Mining, Clustering, Ward's Method, DBSCAN, K-Means*

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“Perbandingan Metode Ward’s, DBSCAN, dan K-Means untuk Klasterisasi Aduan Pelayanan Masyarakat di Dispendukcapil Kota Surabaya”** dapat terselesaikan dengan baik.

Penulis mengucapkan terima kasih kepada Ibu Kartika Maulida Hindrayani, S.Kom., M.Kom., selaku Dosen Pembimbing I dan Bapak Amri Muhaimin, S.Stat., M.Stat., M.S., selaku Dosen Pembimbing 2 yang telah meluangkan waktunya untuk memberikan bimbingan, nasehat serta motivasi kepada penulis. Selain itu, selama penyusunan skripsi penulis juga banyak menerima bantuan dari berbagai pihak. Untuk itu penulis mengucapkan terima kasih kepada:

1. Kepada yang paling tersayang, penulis menyampaikan terima kasih yang sedalam-dalamnya kepada Mama, dan Mas. Terima kasih atas doa yang tidak pernah putus, kasih sayang yang tulus, serta dukungan, baik moral maupun materi, yang selalu menyertai setiap langkah perjalanan hidup penulis. Berkat kesabaran, pengorbanan, dan kepercayaan yang diberikan, penulis dapat bertahan, bangkit, dan akhirnya menyelesaikan skripsi ini dengan sebaik-baiknya. Segala pencapaian ini tentu tidak lepas dari peran kalian. Doa, perhatian, dan semangat yang kalian berikan selalu menjadi kekuatan bagi penulis dalam menghadapi berbagai tantangan, baik selama masa perkuliahan maupun dalam kehidupan sehari-hari. Untuk Mama dan Mas, penulis berharap semoga Allah SWT membalas setiap kebaikan dan pengorbanan yang telah diberikan dengan umur yang panjang, kesehatan, keberkahan, serta kebahagiaan yang berlipat ganda. Terima kasih telah menjadi tempat pulang, tempat bercerita, sekaligus sumber motivasi terbesar bagi penulis untuk terus melangkah dan tidak menyerah.
2. Kepada Papa diatas sana, terima kasih karena pernah hadir dalam hidup penulis. Kehadiran Papa dihidup ini meninggalkan makna yang begitu besar. Setiap nasihat dan pesan yang pernah terucap akan selalu penulis simpan sebagai pengingat, pegangan, dan penyemangat untuk terus menjalani hidup, menjadi

pribadi yang lebih baik, serta menggapai impian-impian penulis. Semoga Papa tenang di sisi-Nya, dan semoga penulis bisa terus melanjutkan perjuangan dengan tetap membawa nama dan doa Papa dalam setiap langkah.

3. Ibu Prof. Dr. Ir. Novirina Hendrasarie, selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Bapak Dr.Eng.Ir.Dwi Arman Prasetya,ST.,MT.,IPU., selaku Koordinator Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur.
5. Dosen-dosen Program Studi Sains Data UPN “Veteran” Jawa Timur yang sudah meluangkan waktu guna berkontribusi pada penelitian ini.
6. Secara khusus, penulis menyampaikan terima kasih kepada Cintya, Debora, dan Domi, sahabat setia sejak masa SMA, atas kebersamaan, kekompakan, serta energi positif yang terus menguatkan hingga saat ini. Walaupun dipisahkan oleh jarak dan kesibukan masing-masing, kalian tetap menjadi pendengar yang baik, pemberi semangat, sekaligus sumber motivasi. Terima kasih atas pertemanan yang tidak lekang oleh waktu. Kalian adalah tempat berbagi cerita tanpa rasa cemas, penyemangat ketika penulis lelah, dan penguat di masa-masa sulit.
7. Kepada teman-teman tercinta penulis, Ica, Wenny, Enzel, Zulfa, Vera, Lisa, dan Elina, penulis mengucapkan terima kasih karena telah menjadi bagian dari perjalanan yang sangat berharga ini. Terima kasih telah bertumbuh bersama dalam setiap fase perkuliahan, melalui hari-hari yang dipenuhi tawa dan duka, berbagai cerita serta berbagai momen yang tak terduga.
8. Teruntuk teman seperjuangan penulis Selenia, Naomi dan seluruh teman-teman lainnya, penulis ucapkan terima kasih sudah menemani proses skripsi dengan kondisi yang terkadang tidak terduga. Terima kasih telah mendengar keluh kesah penulis dan memberi semangat kepada penulis.
9. Rekan-rekan keluarga mahasiswa Prodi Sains Data Universitas Pembangunan Nasional “Veteran” Jawa Timur terima kasih atas dukungan dan kerjasamanya selama menempuh Pendidikan serta penyelesaian penyusunan skripsi ini.
10. Semua pihak yang tidak dapat disebutkan satu-persatu, yang telah banyak membantu memberikan pemikiran demi kelancaran dan keberhasilan penyusunan skripsi.

11. Penulis juga ingin mengucapkan terima kasih kepada diri sendiri karena telah bertahan dan terus berusaha sampai tahap ini. Terima kasih karena tetap melangkah dan mau belajar menerima proses, memperbaiki kekurangan, serta berani bangkit setiap kali merasa tertinggal. Setiap waktu, tenaga, dan air mata yang tercurah menjadi bukti bahwa penulis mampu melewati perjalanan ini dengan penuh keteguhan. Semoga pencapaian ini menjadi pengingat bahwa penulis sanggup menghadapi tantangan berikutnya, terus berkembang, dan tetap percaya pada kemampuan diri.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, 4 Maret 2026

Penulis

DAFTAR ISI

HALAMAN SAMPUL.....	i
LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI.....	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL	xv
DAFTAR LAMPIRAN	xvi
DAFTAR NOTASI.....	xvii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	7
1.3. Batasan Masalah.....	7
1.4. Tujuan Penelitian	8
1.5. Manfaat Penelitian	8
BAB II TINJAUAN PUSTAKA.....	11
2.1. Penelitian Terdahulu	11
2.2. Kerangka Teori.....	17
2.2.1. Dinas Kependudukan dan Pencatatan Sipil (Dispendukcapil).....	17
2.2.2. <i>Text Mining</i>	18
2.2.3. <i>Text Preprocessing</i>	19
2.2.4. Pembobotan Kata (TF-IDF)	23
2.2.5. <i>Sum of Squared Errors</i> (SSE)	24
2.2.6. Analisis Klaster	24
2.2.7. <i>Ward's Method</i>	25
2.2.8. DBSCAN	27
2.2.9. K-Means	28

2.2.10.	<i>Silhouette Coefficient</i>	30
2.2.11.	<i>Index Davies-Bouldin (IDB)</i>	31
2.2.12.	<i>Lexicon Based</i>	32
2.2.13.	<i>Word Cloud</i>	33
2.2.14.	GUI.....	33
2.2.15.	<i>Streamlit</i>	34
BAB III METODOLOGI PENELITIAN.....		35
3.1.	Variabel Penelitian dan Sumber Data.....	35
3.2.	Langkah Analisis	35
3.2.1.	Pengumpulan Data.....	36
3.2.2.	<i>Preprocessing</i> Data	37
3.2.3.	Pembobotan TF-IDF	38
3.2.4.	Pemodelan Data.....	39
3.2.5.	Evaluasi Model.....	42
3.2.6.	Analisis Kluster	43
3.2.7.	Implementasi GUI	44
3.3.	Desain Sistem	44
BAB IV HASIL DAN PEMBAHASAN		51
4.1.	Hasil Penelitian.....	51
4.1.1.	Pengumpulan Data.....	51
4.1.2.	<i>Preprocessing</i> Data	51
4.1.3.	Pembobotan TF-IDF	60
4.1.4.	Pemodelan Data.....	63
4.1.5.	Evaluasi Model.....	94
4.1.6.	Analisis Kluster	99
4.2.	Tampilan Antar Muka (<i>Graphical User Interface</i>).....	107
4.2.1.	Halaman Utama	107
4.2.2.	Halaman <i>Upload</i> Data	108
4.2.3.	Halaman <i>Preprocessing</i>	109
4.2.4.	Halaman TF-IDF	113
4.2.5.	Halaman Pemodelan Ward	114
4.2.6.	Halaman Analisis Emosi	118

BAB V PENUTUP.....	123
5.1. Kesimpulan	123
5.2. Saran Pengembangan	124
DAFTAR PUSTAKA.....	127
LAMPIRAN.....	131

DAFTAR GAMBAR

Gambar 2. 1. Ilustrasi Proses <i>Cleaning</i>	20
Gambar 2. 2. Ilustrasi Proses <i>Tokenizing</i>	21
Gambar 2. 3. Ilustrasi Proses <i>Normalization</i>	21
Gambar 2. 4. Ilustrasi Proses <i>Stopword Removal</i>	22
Gambar 2. 5. Ilustrasi Proses <i>Stemming</i>	22
Gambar 3. 1. Diagram Alir Langkah Analisis	36
Gambar 3. 2. <i>Wireframe</i> Halaman Utama atau Beranda	44
Gambar 3. 3. <i>Wireframe</i> Halaman Proses	45
Gambar 3. 4. Diagram Aktivitas	46
Gambar 4. 1. <i>Output</i> Dendogram	69
Gambar 4. 2. <i>WordCloud</i> Klaster 1 (Ward)	72
Gambar 4. 3. <i>WordCloud</i> Klaster 2 (Ward)	72
Gambar 4. 4. <i>WordCloud Noise</i> (DBSCAN)	79
Gambar 4. 5. <i>WordCloud</i> Klaster 0 (DBSCAN)	80
Gambar 4. 6. <i>WordCloud</i> Klaster 1 (DBSCAN)	81
Gambar 4. 7. <i>WordCloud</i> Klaster 2 (DBSCAN)	81
Gambar 4. 8. <i>WordCloud</i> Klaster 3 (DBSCAN)	82
Gambar 4. 9. <i>WordCloud</i> Klaster 4 (DBSCAN)	82
Gambar 4. 10. <i>WordCloud</i> Klaster 5 (DBSCAN)	83
Gambar 4. 11. Grafik K Optimal K-Means	88
Gambar 4. 12. <i>WordCloud</i> Klaster 0 (K-Means)	92
Gambar 4. 13. <i>WordCloud</i> Klaster 1 (K-Means)	93
Gambar 4. 14. <i>WordCloud</i> Klaster 2 (K-Means)	93
Gambar 4. 15. <i>Pie Chart</i> Emosi	101
Gambar 4. 16. Distribusi Emosi Antar Klaster (Ward)	103
Gambar 4. 17. Distribusi Emosi Antar Klaster (DBSCAN)	104
Gambar 4. 18. Distribusi Emosi Antar Klaster (K-Means)	106
Gambar 4. 19. Tampilan Halaman Utama	108
Gambar 4. 20. Tampilan Halaman <i>Upload</i> data	108
Gambar 4. 21. Tampilan Halaman <i>Upload Data (Preview Data)</i>	109

Gambar 4. 22.	Tampilan Halaman <i>Preprocessing</i> (Data mentah).....	110
Gambar 4. 23.	Tampilan Halaman <i>Preprocessing</i> (<i>Cleaning</i>).....	110
Gambar 4. 24.	Tampilan Halaman <i>Preprocessing</i> (<i>Tokenizing</i>).....	111
Gambar 4. 25.	Tampilan Halaman <i>Preprocessing</i> (<i>Normalization</i>)	111
Gambar 4. 26.	Tampilan Halaman <i>Preprocessing</i> (<i>Stopword</i>).....	112
Gambar 4. 27.	Tampilan Halaman <i>Preprocessing</i> (<i>Stemming</i>)	113
Gambar 4. 28.	Tampilan Halaman TF-IDF.....	114
Gambar 4. 29.	Tampilan Halaman Pemodelan Ward (Dendogram).....	115
Gambar 4. 30.	Tampilan Halaman Pemodelan Ward (K dan Evaluasi)	115
Gambar 4. 31.	Tampilan Halaman Pemodelan Ward (<i>WordCloud</i> Klaster 1-1)116	
Gambar 4. 32.	Tampilan Halaman Pemodelan Ward (<i>WordCloud</i> Klaster 1-2)117	
Gambar 4. 33.	Tampilan Halaman Pemodelan Ward (<i>WordCloud</i> Klaster 2-1)117	
Gambar 4. 34.	Tampilan Halaman Pemodelan Ward (<i>WordCloud</i> Klaster 2-2)117	
Gambar 4. 35.	Tampilan Halaman Analisis Emosi (Tabel Frekuensi)	118
Gambar 4. 36.	Tampilan Halaman Analisis Emosi (<i>Pie chart</i>)	119
Gambar 4. 37.	Tampilan Halaman Analisis Emosi (<i>Bar Chart</i>).....	119
Gambar 4. 38.	Tampilan Halaman Analisis Emosi (<i>WordCloud</i> Klaster 1).....	120
Gambar 4. 39.	Tampilan Halaman Analisis Emosi (<i>WordCloud</i> Klaster 2-1) ..	120
Gambar 4. 40.	Tampilan Halaman Analisis Emosi (<i>WordCloud</i> Klaster 2-2) ..	121
Gambar 4. 41.	Tampilan Halaman Analisis Emosi (<i>WordCloud</i> Klaster 2-3) ..	121

DAFTAR TABEL

Tabel 2. 1. Penelitian Terdahulu.....	11
Tabel 2. 2. Kriteria <i>Silhouette Coefficient</i>	31
Tabel 3. 1. Data Penelitian	35
Tabel 4. 1. Data Teks Aduan Masyarakat	51
Tabel 4. 2. Contoh Hasil Proses <i>Cleaning</i>	54
Tabel 4. 3. Contoh Hasil Proses <i>Tokenizing</i>	55
Tabel 4. 4. Contoh Hasil Proses <i>Normalization</i>	57
Tabel 4. 5. Contoh Hasil Proses <i>Stopword Removal</i>	59
Tabel 4. 6. Contoh Hasil Proses <i>Stemming</i>	60
Tabel 4. 7. Hasil Perhitungan TF-IDF.....	62
Tabel 4. 8. Contoh Dokumen Untuk Perhitungan TF-IDF.....	62
Tabel 4. 9. <i>Output</i> Perhitungan Jarak <i>Euclidean</i>	64
Tabel 4. 10. <i>Output</i> Perhitungan Peningkatan SSE.....	66
Tabel 4. 11. <i>Output</i> Perhitungan Jarak <i>Cosine</i>	74
Tabel 4. 12. <i>Output</i> Titik <i>Core, Border, Noise</i>	77
Tabel 4. 13. <i>Output</i> Pencarian K Optimal Berdasarkan Nilai VRC.....	87
Tabel 4. 14. <i>Output</i> pemodelan K-Means	90
Tabel 4. 15. Evaluasi Model.....	97
Tabel 4. 16. Distribusi Emosi Antar Klaster (Ward).....	102
Tabel 4. 17. Distribusi Emosi Antar Klaster (DBSCAN)	104
Tabel 4. 18. Distribusi Emosi Antar Klaster (K-Means).....	106

DAFTAR LAMPIRAN

Lampiran 1. Dataset	131
Lampiran 2. <i>Source Code</i> Skripsi	132
Lampiran 3. <i>Source Code</i> GUI.....	133

DAFTAR NOTASI

$TF_{t,d}$	: banyaknya kata t pada dokumen d
tf	: banyaknya kata pada dokumen
IDF_t	: nilai <i>inverse document frequency</i>
n	: banyaknya dokumen
df_t	: banyaknya dokumen yang memiliki kata t
$W_{t,d}$	: nilai bobot akhir
$d(i, j)$	: jarak antara data i dan j
x_i	: koordinat x untuk data i
x_j	: koordinat x untuk data j
$\Delta SSE_{A,B}$	: peningkatan nilai SSE jika klaster A dan B digabungkan
n_A, n_B	: jumlah anggota (dokumen) di dalam klaster A dan klaster B
$d(A, B)^2$	: kuadrat jarak Euclidean antara pusat klaster A dan B
x_{ij}	: nilai untuk objek ke- i dalam klaster ke- j
p	: jumlah variabel yang diukur
n	: jumlah objek dalam klaster yang terbentuk
IDB	: <i>Index Davies Bouldin</i>
k	: jumlah klaster
R_{ij}	: rasio antara kompaksi dan separasi antara klaster i dan j
S_i	: kompaksi (<i>intra-cluster distance</i>) klaster i
S_j	: kompaksi (<i>intra-cluster distance</i>) klaster j
D_{ij}	: pemisah (<i>inter-cluster distance</i>) antara pusat klaster i dan j
C_i	: pusat i
C_j	: pusat j
d	: jarak
x_k	: elemen dalam klaster
$S(i)$	: nilai <i>Silhouette Coefficient</i> pada <i>object</i> ke- i
(i)	: <i>object</i> yang akan diteliti
$a(i)$	: rata-rata i jarak i antar anggota i dalam klaster yang sama
$b(i)$	: nilai minimal jarak rata-rata antara item i dengan objek pada klaster lain
V_{ik}	: <i>centroid</i> (rata-rata klaster ke- k untuk dokumen ke-1)

n_k	: banyaknya dokumen yang menjadi anggota klaster ke- k
X_{ji}	: frekuensi kemunculan kata ke- j pada dokumen ke- i
K	: banyak klaster
N	: banyak dokumen
m	: banyak kata
I	: banyak dokumen yang menjadi anggota klaster ke- k
x_{ji}	: frekuensi kemunculan kata $ke-j$ dalam klaster dokumen $ke-i$
y_{jk}	: frekuensi kemunculan kata ke- j pada pusat klaster ke- k
y_{jl}	: frekuensi kemunculan kata ke- j pada pusat klaster ke- l
SSE_k	: nilai SSE pada klaster ke- k
x_i	: dokumen ke- i
C_k	: himpunan anggota klaster ke- k
μ_k	: centroid (pusat klaster) dari klaster ke- k
$\ x_i - \mu_k\ ^2$	: kuadrat jarak Euclidean antara data x_i dan centroid μ_k