



TESIS - SS14 2501

**ANALISIS SENTIMEN PENDAPAT MASYARAKAT
TERHADAP PEMBANGUNAN INFRASTRUKTUR
KOTA SURABAYA MELALUI TWITTER DENGAN
MENGUNAKAN *SUPPORT VECTOR MACHINE*
DAN *NEURAL NETWORK***

ZAKYA REYHANA

NRP. 06211650010010

DOSEN PEMBIMBING

Dr. Kartika Fithriasari, M.Si.

Dr. Raden Mohamad Atok, M.Si.

PROGRAM MAGISTER

JURUSAN STATISTIKA

FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA

INSTITUT TEKNOLOGI SEPULUH NOPEMBER

SURABAYA

2018



THESIS - SS142501

**SENTIMENT ANALYSIS OF CITIZEN OPINION
ABOUT SURABAYA CITY INFRASTRUCTURE'S
DEVELOPMENT ON TWITTER
USING *SUPPORT VECTOR MACHINE* AND
*NEURAL NETWORK***

ZAKYA REYHANA

NRP. 06211650010010

SUPERVISOR

Dr. Kartika Fithriasari, M.Si.

Dr. Raden Mohamad Atok, M.Si.

MASTER PROGRAM

DEPARTMENT OF STATISTICS

FACULTY OF MATHEMATICS, COMPUTING, AND DATA SCIENCE

INSTITUT TEKNOLOGI SEPULUH NOPEMBER

SURABAYA

2018

**ANALISIS SENTIMEN PENDAPAT MASYARAKAT TERHADAP
PEMBANGUNAN INFRASTRUKTUR KOTA SURABAYA MELALUI
TWITTER DENGAN MENGGUNAKAN *SUPPORT VECTOR MACHINE*
DAN *NEURAL NETWORK***

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
Magister Sains (M.Si)

Di

Institut Teknologi Sepuluh Nopember

Oleh:

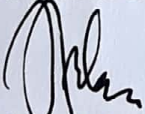
ZAKYA REYHANA
NRP. 06211650010010

Tanggal Ujian : 10 Juli 2018
Periode Wisuda : September 2018

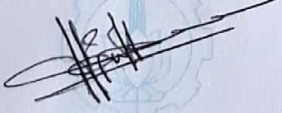
Disetujui Oleh:


1. Dr. Kartika Fithriasari, M.Si.
NIP. 19691212 199303 2 002

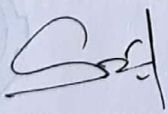
(Pembimbing I)


2. Dr. Raden Mohamad Atok, M.Si.
NIP. 19710915 199702 1 001

(Pembimbing II)


3. Prof. Nur Iriawan, M.Ikomp., Ph.D.
NIP. 19621015 198803 1 002

(Penguji)

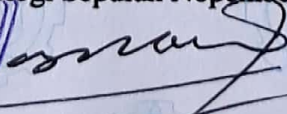

4. Santi Wulan Purnami, M.Si., Ph.D.
NIP. 19720923 199803 2 001

(Penguji)

Dekan

Fakultas Matematika, Komputasi, dan Sains Data
Institut Teknologi Sepuluh Nopember




Prof. Dr. Basuki Widodo, M.Sc.
NIP. 19650605 198903 1 002

**ANALISIS SENTIMEN PENDAPAT MASYARAKAT
TERHADAP PEMBANGUNAN INFRASTRUKTUR KOTA
SURABAYA MELALUI TWITTER DENGAN
MENGUNAKAN *SUPPORT VECTOR MACHINE* DAN
*NEURAL NETWORK***

Nama Mahasiswa : Zakya Reyhana
NRP : 06211650010010
Pembimbing : Dr. Kartika Fithriasari, M.Si.
Co-Pembimbing : Dr. Raden Mohamad Atok, M.Si.

ABSTRAK

Pembangunan infrastruktur Kota Surabaya tengah digiatkan dan dalam prosesnya, berdampak pada aktivitas masyarakat sehari-hari. Berimbas dari dampak, masyarakat kerap memberi komentar atau pendapat masing-masing melalui media apa saja, termasuk media sosial yaitu Twitter. Twitter adalah layanan jejaring sosial dan mikroblog yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter. Pendapat masyarakat tersebut merupakan informasi penting tentang reaksi mereka terhadap program kota yang dilakukan, maka penting untuk didata dan dianalisis. Penelitian yang dilakukan adalah analisis sentimen masyarakat dengan menggunakan metode *Support Vector Machine* dan *Neural Network*. Metode algoritma yang efektif untuk klasifikasi teks diantaranya adalah *Support Vector Machine* (SVM) dan *Neural Network* (NN). Ketepatan klasifikasi SVM sebesar 92,67%, sedangkan NN 80,00%. Untuk perbandingan, metode SVM memberikan akurasi lebih baik daripada NN.

Kata Kunci: Analisis Sentimen, *Neural Network*, Infrastruktur, Klasifikasi, *Support Vector Machine*, Surabaya, Twitter

(Halaman ini sengaja dikosongkan)

SENTIMENT ANALYSIS OF CITIZEN OPINION ABOUT SURABAYA CITY INFRASTRUCTURE'S DEVELOPMENT ON TWITTER USING *SUPPORT VECTOR MACHINE* AND *NEURAL NETWORK*

Name : Zakya Reyhana
Student Identity Number : 06211650010010
Supervisor : Dr. Kartika Fithriasari, M.Si.
Co-Supervisor : Dr. Raden Mohamad Atok, M.Si.

ABSTRACT

The development of Surabaya's infrastructure was being intensified and in the process, it affects the daily activities of the community. This event had an impact on public comments or opinions about it through any media, including social media such as Twitter. Twitter was a social networking service and a microblog that allows its users to send and read text-based messages of up to 140 characters. The public opinion was an important information about their reaction to the city's development program, so it was important to be recorded and analyzed. The research was about citizen sentiment analysis using Support Vector Machine and Neural Network methods. Effective method algorithms for text classification were Support Vector Machine (SVM) and Neural Network (NN). The accuracy of SVM classification was 92.67%, while NN 80.00%. For comparison, the SVM method gives better accuracy than NN.

Keywords: Sentiment Analysis, Neural Network, Infrastructure, Classification, Support Vector *Machine*, Surabaya, Twitter

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Syukur Alhamdulillah, penulis panjatkan kepada Allah SWT karena atas limpahan karunia, berkat, dan nikmat diberikan untuk penulis sehingga laporan tesis yang berjudul “**Analisis Sentimen Pendapat Masyarakat Terhadap Pembangunan Infrastruktur Kota Surabaya Melalui Twitter dengan Menggunakan *Support Vector Machine* dan *Neural Network***” ini dapat diselesaikan sesuai dengan waktu yang ditargetkan. Buku tesis ini disusun guna memenuhi salah satu syarat kelulusan untuk meraih gelar Magister Sains pada Program Pasca Sarjana Statistika di Institut Teknologi Sepuluh Nopember (ITS) Surabaya.

Dalam proses penulisan laporan tesis ini tidak terlepas dari dukungan dari berbagai pihak. Maka dari itu penulis ingin mengucapkan banyak terima kasih kepada:

1. Orang tua penulis, Bapak Bambang Moelyono dan Ibu Ramlah Maricar, juga kakak kandung penulis Muhammad Taufiq Ramadhan, atas segala dukungan, arahan, pengertian, kesabaran dan doa kebaikan yang tidak pernah putus untuk penulis.
2. Ibu Dr. Kartika Fithriasari, M.Si selaku dosen wali dan juga dosen pembimbing yang telah mengarahkan dan membimbing penulis selama proses kuliah hingga akhir penyusunan tesis.
3. Bapak Dr. Raden Mohamad Atok, M.Si, selaku pembimbing tugas akhir tesis yang memberikan bimbingan, arahan dan motivasi dalam penyusunan buku tesis.
4. Bapak Prof. Nur Iriawan, M.Ikomp., Ph.D dan Ibu Santi Wulan Purnami, M.Si., Ph.D, selaku dosen penguji yang telah mengoreksi kesalahan-kesalahan yang pernah ada, mengarahkan dan memberikan saran yang membangun dalam proses penyusunan buku tesis.
5. Bapak Dr. Suhartono, M.Sc, selaku Ketua Jurusan Statistika ITS, Bapak Dr. Rer.pol Heri Kuswanto, M.Si, selaku Ketua Prodi Pasca Sarjana Statistika ITS,

beserta jajarannya, atas semua fasilitas, arahan, inspirasi, dan motivasi yang diberikan selama masa studi.

6. Bapak dan Ibu dosen Statistika ITS yang telah mengajarkan ilmu dan pengalaman mereka, membimbing dan memotivasi selama proses studi berlangsung.
7. Saidah Zahrotul Jannah, sahabat dan teman sebimbingan, teman diskusi dari mencari topik tesis sampai menyusun hingga akhir, dengan sabar menjawab seluruh pertanyaan, memberikan informasi, dan dukungan, dan perhatian yang lebih dari cukup.
8. Arlene Henny Hiariy, Sella Aji Oktarin, dan Milasari, selaku sahabat seperjuangan. Meski beda pembimbing, beda metode, dan beda topik, tetapi selalu memberikan informasi-informasi, motivasi, dan saling memberikan dukungan.
9. Kiki Ferawati, Tri Murniati, dan T.Dwi Ary W., serta teman-teman seperjuangan selama kuliah di Program Magister Jurusan Statistika ITS atas dukungan, berbagi pengalaman dan kebaikan-kebaikan lainnya.
10. Sahabat-sahabat di luar lingkungan kampus, Rucita N. D. L., Anastrinia S.H., Bellafaatiha, Indah K.W., Risma H., Diana S.A., Aviana G.L., dll, keluarga besar serta semua pihak yang berperan dan membantu dari awal hingga penyusunan tesis ini telah usai.

Penulis berharap agar tesis ini bermanfaat dan dapat menambah wawasan keilmuan. Saran dan kritik yang membangun dari semua pihak demi perbaikan lebih lanjut akan sangat diterima.

Surabaya, Juli 2018

Penulis

DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN	iii
ABSTRAK	v
ABSTRACT.....	vii
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN.....	xvii
BAB 1	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	4
1.5 Batasan Masalah.....	4
BAB 2	5
2.1 Analisis Sentimen Pada Text Mining.....	5
2.2 Preprocessing Data.....	6
2.3 Twitter	8
2.4 Penelitian Sebelumnya	8
2.5 Support Vector Machine	10
2.5.1 SVM <i>Linierly Separable</i>	11
2.5.2 SVM <i>Linear Nonseparable</i>	15
2.5.3 SVM <i>Nonlinear Separable</i>	17
2.6 Neural Network	20
2.7 Cross Validation	23
2.8 Teknik Evaluasi Kinerja Model Klasifikasi	24
2.9 Pembentukan Klaster dengan Metode K-Means.....	26
2.10 Infrastruktur Surabaya.....	29
BAB 3	33

3.1	Data dan Sumber Data	33
3.2	Struktur Data.....	33
3.3	Langkah Analisis	35
3.4	Diagram Alir.....	37
BAB 4.....		41
4.1	Klasifikasi Sentimen Data Twitter Menggunakan Metode SVM dan Neural Network.	41
4.1.1	Karakteristik Data	41
4.1.2	Perangkat Pengujian	44
4.1.3	Preprocessing Data	44
4.1.4	Document Term Matrix	49
4.1.5	Term-Weighting TF-IDF	50
4.1.6	Klasifikasi dengan SVM.....	51
4.1.7	Klasifikasi dengan Neural Network.....	55
4.1.8	Perbandingan Metode Klasifikasi.....	57
4.2	Pembentukan Kluster	58
4.2.1	Pembentukan Kluster Sentimen Negatif.....	59
4.2.2	Pembentukan Kluster Sentimen Positif.....	62
BAB 5.....		65
5.1	KESIMPULAN.....	65
5.2	SARAN.....	65
DAFTAR PUSTAKA.....		67
LAMPIRAN		73
BIOGRAFI PENULIS		99

DAFTAR TABEL

Tabel 3.1. Struktur Data	34
Tabel 3.2. Contoh Bentuk Data	34
Tabel 4.1. Contoh Data Awal	42
Tabel 4.2. Statistika Deskriptif	43
Tabel 4.3. Spesifikasi <i>Hardware</i> dan <i>Software</i>	44
Tabel 4.4. Document Term Matrix 110 Kata	50
Tabel 4.5. Matriks DTM dengan pembobotan TF-IDF	51
Tabel 4.6. Ketepatan Klasifikasi Metode SVM Non-linier Kernel RBF $C=0,01$, Gamma=1000	54
Tabel 4.7. Ringkasan Struktur Model Neural Network	56
Tabel 4.8. Model Neural Network di Tiap Neuron dalam 1 Hidden Layer	56
Tabel 4.9. Model Neural Network untuk Klasifikasi <i>Tweet</i> Sentimen Negatif dan Positif	57
Tabel 4.10. Ketepatan Klasifikasi Metode NN 1 Hidden Layer 71 Neuron	57
Tabel 4.11. Perbandingan Ketepatan Klasifikasi SVM dan NN	58
Tabel 4.12. Frekuensi Term	59
Tabel 4.13. Nilai Pseudo F Jumlah Klaster Sentimen Negatif	61
Tabel 4.14. Nilai Pseudo F Jumlah Klaster Sentimen Positif	64

(Halaman ini sengaja dikosongkan)

DAFTAR GAMBAR

Gambar 2.1.	(a) SVM <i>Linear Separable</i> ; (b) SVM <i>Linear Nonseparable</i> ; (c) SVM <i>Nonlinear Separable</i> (Haerdle, Prastyo, & Hafner, 2014). .	11
Gambar 2.2.	Bidang pemisah terbaik dengan margin (m) terbesar (Haerdle, Prastyo, & Hafner, 2014).....	12
Gambar 2.3.	Bidang pemisah terbaik dengan margin terbesar linier <i>non separable</i> (Haerdle, Prastyo, & Hafner, 2014).....	15
Gambar 2.4.	Mapping dari Dua Dimensi Data <i>Space</i> (Kiri) ke Tiga Dimensi <i>Feature Space</i> (Kanan) (Haerdle, Prastyo, & Hafner, 2014)	17
Gambar 2.5.	Arsitektur <i>Backpropagation</i>	21
Gambar 2.6.	<i>Confusion Matrix</i> dua kelas klasifikasi	26
Gambar 3.1.	Diagram Alir Penelitian.....	37
Gambar 4.1.	<i>Pie Chart</i> Data <i>Tweet</i> Berdasarkan Kandungan Sentimen.....	42
Gambar 4.2.	Histogram Frekuensi Data Awal	43
Gambar 4.3.	Distribusi Panjang <i>Tweet</i> dan Label Kelasnya	43
Gambar 4.4.	Histogram Frekuensi Data Setelah Praproses.....	49
Gambar 4.5.	Scatter Plot Variabel Listrik dan Nyala.....	52
Gambar 4.6.	Plot Pencarian Parameter C dan Sigma dengan Teknik <i>Grid Search</i>	53
Gambar 4.7.	Wordcloud Term Sentimen Negatif	60
Gambar 4.8.	Elbow Klaster Sentimen Negatif	60
Gambar 4.9.	Wordcloud Terms Sentimen Positif	62
Gambar 4.10.	Elbow Klaster untuk Sentimen Positif	63

(Halaman ini sengaja dikosongkan)

DAFTAR LAMPIRAN

Lampiran 1.	Istilah total sebanyak 2.824 istilah pada 1500 <i>tweet</i>	73
Lampiran 2.	Pemindaian Grid dengan teknik Stratified Cross Validation	82
Lampiran 3.	Keterangan Model SVM.....	83
Lampiran 4.	Prediksi data <i>testing</i>	87
Lampiran 5.	Training Model Neural Network	88
Lampiran 6.	Bobot Neural Network untuk Neuron ke-1 di dalam hidden layer.	90
Lampiran 7.	Sepuluh klaster yang dibangun berdasarkan kandungan sentimen negatif	92
Lampiran 8.	Sepuluh klaster yang dibangun berdasarkan kandungan sentimen positif.....	94
Lampiran 9.	Syntax for Crawling Data.....	95
Lampiran 10.	<i>Preprocessing Data</i>	96
Lampiran 11.	<i>Syntax</i> Membentuk Model SVM.....	97
Lampiran 12.	<i>Syntax</i> Membentuk Model NN.....	98

(Halaman ini sengaja dikosongkan)

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Kota Surabaya merupakan kota nomor dua terbesar di Indonesia, setelah Ibu Kota Jakarta. Masalah yang dihadapi kota-kota besar yang paling sering terjadi ialah kemacetan. Disampaikan walikota, untuk tahun 2017, Pemerintah kota akan memfokuskan proyek infrastruktur. Ibu Tri Rismaharini menyebutkan bahwa pemerintah kota akan mendahulukan usulan terkait aksesibilitas untuk infrastruktur, lalu ekonomi, pendidikan, dan kesehatan. Alasannya karena di situlah ketangguhan kota untuk menghadapi persaingan global akan teruji (Surabaya, 2017). Kepala Badan Perencanaan dan Pembangunan Kota (Bapekko) Surabaya Agus Iman Sohanji menyatakan bahwa visi utama pembangunan Kota Surabaya adalah meningkatkan daya saing ekonomi lokal Kota Surabaya. Salah satu penyebab mobilitas perkotaan yang kurang lancar adalah terurainya kemacetan di berbagai wilayah kota. Bila jalanan lancar, berbagai kepentingan warga tidak akan tersendat (Prasetyo, 2017).

Menteri Koordinator Perekonomian, Darmin Nasution, mengatakan Kota Surabaya sebagai contoh konsep kota pintar atau *smart city* yang sudah dikembangkan di Indonesia (Oktara, 2017). Definisi kota pintar (*smart city*) merupakan pengembangan, penerapan, dan implementasi teknologi digital (TIK) yang diaplikasikan pada sebuah wilayah / kota. Kota pintar mampu meningkatkan kualitas kehidupan, mengurangi biaya dan sumber konsumsi, serta membantu meningkatkan interaksi antar kota dan warganya secara efektif. Kota pintar dianggap mampu menyerap dan menganalisa informasi dengan baik dan cepat (Kristalina, 2015). Implementasi teknologi digital seperti aplikasi media sosial memperluas pandangan teknologi informasi dan komunikasi pada sektor publik dan digunakan untuk meningkatkan transparansi, partisipasi, dan kolaborasi pemerintah di pemerintahan pusat Amerika Serikat. Keberhasilan, dampak, dan kinerja dari bentuk dua arah dan jaringan interaksi dapat memberikan wawasan untuk memahami kepatuhan terhadap mandat Inisiatif Pemerintahan Terbuka atau *Open*

Government Initiative (OGI) (Mergel, 2013). Salah satu media untuk berbagi informasi mengenai keadaan Kota Surabaya untuk pemerintah dan antar warga asli maupun pendatang Surabaya melalui media sosial yaitu Twitter. Informasi yang ditemukan di Twitter selain berupa pelaporan atau keluhan, ada juga yang berupa apresiasi suatu hasil kerja atau kebijakan. Informasi positif, negatif, atau dan juga netral ini adalah sentimen masyarakat yang layak dianalisis agar pantas dijadikan acuan pengukuran kinerja kota.

Twitter adalah layanan jejaring sosial dan mikroblog yang memungkinkan pengguna untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter, yang dikenal dengan sebutan kicauan (*tweet*) (Wikipedia, 2017) . Twitter memperkenankan pengguna mengirim *tweet* mengenai apa yang mereka lakukan atau memberi tautan ke website lain yang diminati. Orang bisa memilih untuk “mengikuti” *tweet* pengguna terpilih dan dapat me-*retweet* atau mem-*posting* ulang *tweet* seseorang untuk dilihat orang lain (Ezunumah, 2013). Penggunaan analisis sentimen pada situs Twitter bertujuan untuk mengklasifikasikan konten secara implisit yang dituliskan oleh pengguna Twitter.

Analisis sentimen merupakan suatu deteksi berbasis komputasi dan pembelajaran terhadap pendapat atau pandangan (sentimen), emosi, dan subjektivitas di dalam teks. Sebagai aplikasi *text mining* yang spesial, analisis sentimen berhubungan dengan ekstraksi otomatis opini positif atau negatif dari teks (He, Wu, Yan, Akula, & Shen, 2015). Data *tweet* diklasifikasikan menjadi opini positif dan negatif dengan menggunakan metode klasifikasi teks. Metode yang paling sering digunakan dalam klasifikasi teks diantaranya adalah *Support Vector Machine* (SVM) karena metode ini sangat cepat dan efektif (Feldman & Sanger, 2007). SVM adalah teknik yang sesuai untuk kasus klasifikasi positif negatif (biner), yang terkait dan mengandung unsur statistika terapan non-parametrik, jaringan saraf tiruan, dan *machine learning* (ML) (Auria & Moro, 2008). Selain SVM, metode lain yang bisa digunakan ialah *Neural Network*. *Neural Network* dapat digunakan untuk memodelkan hubungan yang kompleks antara input dan output untuk menemukan pola-pola pada data. Metode *Neural Network* telah banyak diaplikasikan di berbagai bidang, diantaranya: *Time Series* (Fithriasari, dkk, 2013), regresi (Bataneh & Marler, 2017), dan klasifikasi (Mandal & Banerjee,

2015). Menurut Habibi (2016), nilai persentase evaluasi performa yang dihasilkan di dalam penelitiannya menunjukkan bahwa klasifikasi dari analisis sentimen dengan metode *neural network* ber algoritma *backpropagation* memberikan hasil yang baik.

Berdasarkan kandungan sentimennya, kumpulan teks tersebut dapat diolah dan dilakukan pengelompokan teks untuk mendapatkan topik spesifik dari laporan masyarakat. Dhini, dkk (2017) di dalam penelitiannya memperoleh hasil pengelompokan teks yang menunjukkan bahwa mayoritas keluhan publik terkait dengan irigasi dan perbaikan jalan.

Penelitian sebelumnya mengenai analisis sentimen dilakukan oleh Taufik Kurniawan (2017) yang meneliti sentimen pengguna Twitter terhadap media mainstream TV One, Metro TV, dan Kompas TV menggunakan metode *Naive Bayes Classifier* dan *Support Vector Machine*. Selain itu, penelitian mengenai analisis sentimen juga dilakukan oleh Bo Pang (2002), Ni Wayan Sumartini Sarawaswati (2013), Heru Susanto (2014), Nuvirta Monarizqa (2014), Novia Dwi Putranti (2014), Syahfitri Kartika Lidya (2015), dan Habibi (2016).

1.2 Rumusan Masalah

Penelitian klasifikasi data *tweet* menggunakan metode SVM untuk mengklasifikasi opini negatif dan positif telah banyak dilakukan. Namun, penelitian-penelitian tersebut lebih fokus membahas sentimen di dalam perusahaan atau partai politik tertentu. Padahal analisis sentimen juga bisa dilakukan terhadap topik yang lain. Adapun metode yang digunakan adalah metode yang tradisional. Maka, pada penelitian ini dirumuskan permasalahan, bagaimana analisis sentimen terhadap data Twitter mengenai pembangunan infrastruktur Kota Surabaya dengan menggunakan metode SVM dan Neural Network? Bagaimana cara mengetahui topik kata spesifik apa saja yang sering disebutkan oleh masyarakat Surabaya terkait pembangunan infrastruktur?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan yang ingin dicapai dalam penelitian ini adalah:

- a. Memperoleh hasil analisis sentimen terhadap data Twitter mengenai pembangunan infrastruktur Kota Surabaya dengan menggunakan metode SVM dan Neural Network.
- b. Mendapatkan topik spesifik yang dibahas dari *tweet* masyarakat Surabaya dengan cara membentuk kluster berdasarkan masing-masing sentimen negatif dan positif.

1.4 Manfaat Penelitian

Manfaat yang ingin dicapai dari hasil penelitian ini adalah :

- a. Memberikan informasi mengenai analisis sentimen dan terhadap data *tweet* dan bagaimana pengklasifikasiannya.
- b. Memberikan gambaran deskriptif kluster topik yang spesifik dari *tweet* masyarakat Surabaya berdasarkan klasifikasi negatif dan positif yang sudah terbentuk.
- c. Sebagai bahan pertimbangan bagi pemerintah Kota Surabaya dan dinas atau instansi yang terkait dengan infrastruktur dalam pengambilan keputusan maupun merumuskan kebijakan/program yang tepat dalam rangka peningkatan kualitas infrastruktur, sumber daya manusia, peningkatan kinerja kota, dan peningkatan kualitas kota.

1.5 Batasan Masalah

Pada penelitian ini, masalah dibatasi pada:

- a. Data berbahasa Indonesia.
- b. Data *tweet* media sosial Twitter di dalam lingkup Surabaya yang dikumpulkan sejak 29 Agustus hingga Oktober 2017 dan Januari hingga 7 Februari 2018.
- c. Sentimen yang diklasifikasikan adalah yang bernilai positif dan negatif secara manual.
- d. Data yang diambil dicari dengan kata kunci: sapawargaSby dan e100ss, akan digabung menjadi satu dataset.

BAB 2

TINJAUAN PUSTAKA

2.1 Analisis Sentimen Pada Text Mining

Analisis teks atau bisa juga disebut sebagai penggalian data (*text mining*) merupakan metodologi dan sebuah proses yang diiringi dengan perolehan kualitas dan informasi yang dapat ditindaklanjuti serta wawasan dari data tekstual. Proses ini melibatkan penggunaan Pemrosesan Bahasa Alami (*Neuro-Linguistic Programming*) atau NLP. Proses penggunaan NLP, penarikan informasi, dan teknik *machine learning* digunakan untuk mengurai data teks tidak terstruktur ke dalam bentuk lebih terstruktur dan mengambil pola dan wawasan dari data tersebut yang berguna oleh pengguna (Sarkar, 2016).

Text mining merupakan penambangan data berupa teks yang diperoleh dari data tidak terstruktur berupa kalimat-kalimat didalam dokumen, lalu dari dokumen tersebut dicari kata-kata yang dapat mewakili isi dokumen untuk dapat menganalisis inti dari dokumen tersebut. Penjelasan lebih sederhananya, *text mining* adalah proses penyaringan wawasan yang dapat ditindaklanjuti atau dianalisa dari teks. Secara spesifik, *text mining* dapat digunakan untuk mengidentifikasi kabar dari media sosial yang dapat ditindaklanjuti untuk organisasi layanan pelanggan. *Text mining* mewakili kemampuan untuk mengambil bahasa tidak terstruktur dalam jumlah yang besar dan cepat dalam mengekstrak wawasan yang berguna untuk pengambilan keputusan. Hal-hal tersebut dilakukan tanpa memaksa seseorang untuk membaca seluruh badan teks (Kwartler, 2017).

Tujuan *text mining* adalah untuk mendeskripsikan dan mengeksplor data tekstual, untuk mengungkap ciri struktural dan memproses prediksi (Lebart, 1998). Perangkat lunak *text mining* dapat digunakan untuk membuat arsip informasi berukuran besar tentang orang atau peristiwa tertentu. Secara umum, *text mining* terbagi menjadi dua tipe, yang pertama disebut “*bag of words*” dan yang satunya lagi disebut “*syntactic parsing*”, dengan kelebihan dan kekurangan masing-masing.

Analisis sentimen, juga disebut *opinion mining*, adalah bidang studi yang menganalisa pendapat, sentimen, penilaian, sikap, dan emosi orang terhadap entitas dan diungkapkan dalam teks tertulis. Entitas dari analisis sentimen dapat berupa

produk, layanan, organisasi, individu, acara, isu, atau topik (Liu, 2015). Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen atau kalimat. Sentimen yang dimaksud bisa berupa negatif, positif, netral, dan lain-lain.

Sentimen negatif, positif, atau netral pada umumnya berupa sikap dalam bentuk verbal atau pendapat yang diungkapkan dalam teks terhadap subjek tertentu. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian ataupun aplikasi mengenai analisis sentimen berkembang pesat, bahkan menurut (Liu, 2012), di Amerika kurang lebih 20-30 perusahaan yang memfokuskan pada layanan analisis sentimen.

2.2 Preprocessing Data

Mengerjakan analisis teks seringkali lebih melibatkan proses daripada analisis statistika atau *machine learning*. Semua algoritma *machine learning*, teknik *supervised* atau *unsupervised*, biasanya diawali dengan *preprocessing* sebelum mengolah dan menganalisis data. Tujuan dilakukannya *preprocessing* ialah untuk menghilangkan *noise*, menyeragamkan bentuk kata dan mengurangi volume kata. Pada tahap *preprocessing* terdiri proses *cleaning*, *case folding*, *parsing*, dan *filtering* (Putranti & Winarko, 2014).

- a. *Cleaning* adalah proses untuk membersihkan dokumen dan kata-kata yang tidak diperlukan untuk mengurangi *noise* pada proses klasifikasi.
- b. *Case folding* adalah proses penyeragaman bentuk huruf serta penghilangan tanda baca. Dalam hal ini, seragam hanya huruf latin dari a sampai z saja.
- c. *Parsing* yaitu proses memecah dokumen menjadi sebuah kata dengan melakukan analisis terhadap kumpulan kata dengan memisahkan kata tersebut dan menentukan struktur sintaksis dari tiap kata tersebut.
- d. Filter bahasa adalah proses untuk memilih *tweet* yang berbahasa Indonesia saja dan jika ditemui kata berbahasa Indonesia tidak baku maka diganti dengan sinonimnya berupa kata baku yang sesuai dengan Kamus Besar Bahasa Indonesia.

Praproses dalam *text mining* cukup rumit karena dalam Bahasa Indonesia terdapat berbagai aturan penulisan kalimat maupun pembentukan kata berimbuhan.

Terdapat empat aturan pembentukan kata berimbuhan (afiks) untuk merubah makna kata dasar yaitu sebagai berikut.

- a. Awalan (prefiks), imbuhan yang dapat ditambahkan pada awal kata dasar. Imbuhan ini terbagi dalam dua jenis.
 - i. Standar, yang mencakup imbuhan ‘di-’, ‘ke-’, dan ‘se-’.
 - ii. Kompleks, yang mencakup imbuhan ‘me-’, ‘be-’, ‘pe-’, dan ‘te-’.Perbedaan antara kedua jenis imbuhan awalan tersebut yaitu penambahan imbuhan awalan standar pada suatu kata dasar tidak merubah kata dasar tersebut, sedangkan imbuhan awalan kompleks pada suatu kata dasar dapat mengubah struktur kata dasar tersebut.
- b. Akhiran (sufiks), imbuhan yang ditambahkan di belakang kata dasar. Sufiks yang sering digunakan yaitu: ‘-i’, ‘-kan’, dan ‘-an’. Selain itu, imbuhan kata yang menunjukkan keterangan kepemilikan seperti: ‘-ku’, ‘-mu’, dan ‘-nya’ serta partikel ‘-lah’, ‘-kah’, ‘-tah’, dan ‘-pun’ juga dapat dikategorikan sebagai sufiks.
- c. Awalan dan akhiran (konfiks), imbuhan yang ditambahkan di depan dan belakang kata dasar (prefiks dan sufiks) secara bersama-sama.
- d. Sisipan (infiks), imbuhan yang ditambahkan di tengah kata dasar.

Proses penarikan informasi (*information retrieval*) berawal ketika pengguna memasukkan sebuah kueri ke dalam sistem. Kueri adalah kalimat-kalimat formal informasi. Dalam hal ini, kueri tidak mengidentifikasi secara unik suatu objek dalam *corpus*. Gaya kueri ini, seringkali hanya dilihat sebagai kumpulan kata. Mekanisme penilaian yang masuk akal yaitu menghitung skor yang merupakan jumlah dari kueri pada antara istilah (*term*) pada kueri dan dokumen. Penetapan bobot istilah dalam dokumen masing-masing yang bergantung pada jumlah kemunculan istilah pada dokumen dilakukan pada mekanisme penilaian. Skor yang dihitung ialah istilah kueri t dan dokumen d berdasarkan bobot t di d . Skema pembobotan ini disebut sebagai frekuensi istilah (*term frequency*) dan dilambangkan dengan $tf_{t,d}$ (Manning & Raghavan, 2008). Frekuensi istilah dapat diperoleh dengan formula pada (2.1).

$$tf_{t,d} = \frac{\text{Jumlah istilah } t \text{ muncul dalam dokumen}}{\text{Total jumlah istilah dalam dokumen}} \quad (2.1)$$

Frekuensi istilah mentah seperti di atas memiliki masalah kritis: semua istilah dianggap sama pentingnya ketika menyangkut nilai relevansi pada suatu kueri. Bahkan, istilah-istilah tertentu memiliki sedikit atau tidak memiliki kekuatan yang membedakan dalam menentukan relevansi. Oleh karena itu, mekanisme untuk melemahkan efek tersebut dengan cara menurunkan bobot dengan frekuensi tinggi, yang didefinisikan sebagai jumlah total kemunculan suatu istilah dalam *corpus*. Identya ialah untuk mengurangi bobot suatu istilah dengan faktor yang tumbuh dengan frekuensi pengumpulannya. Pengukuran bobot pada frekuensi dokumen (*document frequency*) pada suatu istilah dilakukan dengan cara penandaan (*denoting*). Penandaan tersebut didefinisikan sebagai frekuensi dokumen terbalik atau *inverse document frequency* (*idf*) dari suatu istilah term t (Manning & Raghavan, 2008), seperti pada formula (2.2).

$$idf_t = \frac{N}{df_t} \quad (2.2)$$

2.3 Twitter

Twitter adalah salah satu layanan media sosial yang paling populer di internet, membiarkan percakapan berisi “kicauan” individual diantara kumpulan banyak orang. Pencipta Twitter ialah Jack Dorsey di tengah tahun 2006 dengan link URL <http://www.twitter.com>. Twitter mengakui kekuatan dari data historisnya untuk memungkinkan penelitian tentang garis waktu / lini masa perubahan dalam karakteristik yang disimpulkan dari data mereka, seperti, “*mood* (suasana hati)”. Aplikasi *microblog* tersebut memiliki aplikasi pemrograman antarmuka (*Application Programming Interface*) agar developer dapat mengembangkan aplikasi tersebut sesuai dengan keinginan dan kebutuhan masing-masing. Adapun dokumentasi perihal Twitter API dapat dilihat di <http://dev.twitter.com>.

2.4 Penelitian Sebelumnya

Analisis teks terutama analisis sentimen sudah sering dilakukan oleh beberapa peneliti baik luar maupun dalam negeri. Penelitian sebelumnya mengenai analisis sentimen di antaranya yakni:

1. Implementasi *Text Mining* pada Analisis Sentimen Pengguna Twitter Terhadap Media *Mainstream* menggunakan *Naive Bayes Classifier* dan *Support Vector*

Machine oleh Taufik Kurniawan di tahun 2017. Secara keseluruhan perbandingan performa metode NBC dan SVM menunjukkan hasil bahwa performa SVM lebih baik dalam mengklasifikasikan data.

2. Identifikasi Topik Informasi Publik Media Sosial di Kota Surabaya Berdasarkan Klasterisasi Teks Pada Twitter dengan Menggunakan Algoritma K-Means oleh Moh. Hasan Basri pada tahun 2016. Hasil identifikasi topik untuk data keseluruhan didapatkan pelabelan 1, 2, dan 3 berturut-turut memiliki hasil klasifikasi optimal menggunakan algoritma SVM Kernel ‘Linear’ dengan akurasi 95,92%, 95,51%, dan 96,79%.
3. Analisis Sentimen Pada Twitter Mahasiswa Menggunakan Metode *Backpropagation* oleh Robet Habibi pada tahun 2016. Nilai persentase evaluasi performa yang dihasilkan menunjukkan bahwa klasifikasi dari analisis sentimen memberikan hasil yang baik.
4. Visualisasi Data Teks *Twitter* Berbasis Bahasa Indonesia Menggunakan Teknik Pengklasteran oleh Heru Susanto di tahun 2014. Algoritma pengklasteran yang digunakan adalah *K-Means*, *Cascade K-Means* dan *Self-Organizing Map Kohonen*. Hasil yang didapat menunjukkan bahwa *Cascade K-Means* mampu menghasilkan nilai konvergensi kelompok terkecil SSE sebesar 7.073 dan *Dunn Index* sebesar 0,67 dengan distribusi sentimen positif berjumlah 26.332 *tweet*, negatif berjumlah 7.912 *tweet*, dan netral berjumlah 23.050 *tweet*.
5. Penerapan Analisis Sentimen Pada Twitter Berbahasa Indonesia Sebagai Pemberi *Rating* oleh Nuvirta Monarizqa pada tahun 2014. Pada penelitian ini terdapat dua tahap, pertama membentuk *core* aplikasi berupa konfigurasi metode analisis sentimen yang dikembangkan menggunakan pustaka Lucene sebagai alat *stemming* dan algoritma *Support Vector Machine* (SVM), dan tahap kedua adalah membuat aplikasi analisis sentimen dengan menggunakan Java. Hasil yang didapat ialah *core* aplikasi mampu memberikan akurasi sejumlah 73,43% dengan *7-fold cross validation*. Sedangkan untuk aplikasinya mampu memberikan *rating* 4,4 dan akurasi sejumlah 68% untuk kata kunci “Jokowi”, *rating* 2,7 dan akurasi sejumlah 56% untuk kata kunci “Prabowo”,

rating 4,6 dan akurasi sejumlah 70% untuk kata kunci “kalimilk” serta *rating* 3,7 dan akurasi sejumlah 745 untuk kata kunci “sunmor”.

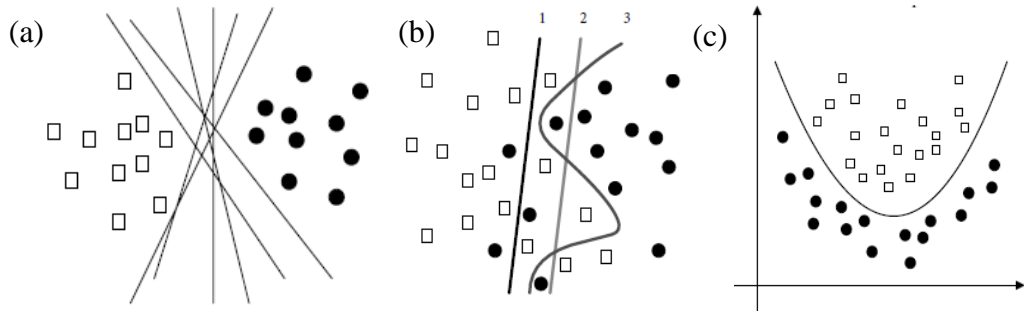
2.5 Support Vector Machine

Algoritma klasifikasi merupakan algoritma *machine learning* terbimbing (*supervised*) yang digunakan untuk mengklasifikasi, mengategorikan, atau melabeli titik data berdasarkan apa yang telah diobservasi sebelumnya. Ada banyak tipe dari algoritma klasifikasi, tetapi algoritma yang cukup efektif untuk klasifikasi teks diantaranya adalah *Support Vector Machine* (SVM) dan *Naive Bayes*. Selain kedua algoritma tersebut, beberapa algoritma lain yang diketahui yaitu regresi logistik, *decision tree*, dan *neural network* (Sarkar, 2016).

Pada *machine learning*, SVM adalah algoritma pembelajaran terbimbing (*supervised*) digunakan untuk klasifikasi, regresi, dan anomali atau pendeteksian pencilan (*outlier*). SVM berusaha menemukan garis pemisah (*hyperplane*) yang terbaik pada ruang input *space*. Prinsip dasar SVM adalah *linear classifier*, mengingat masalah klasifikasi biner, jika kita memiliki data latihan yang tiap titik data atau observasi tergolong pada kelas yang spesifik, algoritma SVM bisa dilatih berdasarkan data tersebut dan bisa menetapkan titik data yang akan datang ke dalam salah satu dari dua kelas kategori. Namun, SVM juga bisa melakukan klasifikasi non-linier dengan pendekatan menarik yang dikenal dengan sebutan trik kernel, dimana fungsi kernel digunakan untuk mengoperasikan ruang fitur berdimensi tinggi yang terpisah secara non-linier. Algoritma SVM mengambil satu set titik data latihan dan mendirikan sebuah *hyperplane* dari koleksi *hyperplane* untuk ruang fitur berdimensi tinggi. Semakin besar batas *hyperplane*, semakin bagus pemisahannya, dan dengan demikian hal ini akan mengurangi kesalahan generalisasi alat yang mengklasifikasi (*classifier*) (Sarkar, 2016).

Garis pemisah (*hyperplane*) terbaik antara kedua kelas didapatkan dengan cara mengukur *margin hyperplane* tersebut dan mencari titik maksimalnya. *Margin* adalah jarak antara *hyperplane* tersebut dengan titik terdekat dari masing-masing kelas. Titik yang terpotong oleh *supporting hyperplane* disebut *support vector*.

Seperti diilustrasikan pada Gambar 2.1, pada klasifikasi, SVM dapat dibedakan menjadi tiga yaitu (a) SVM *linear separable*, (b) SVM *linear nonseparable*, dan (c) SVM *nonlinear separable*.



Gambar 2.1 (a) SVM *Linear Separable* ; (b) SVM *Linear Nonseparable*; (c) SVM *Nonlinear Separable* (Haerdle, Prastyo, & Hafner, 2014).

2.5.1 SVM *Linierly Separable*

Haerdle, dkk (2014) menyatakan, bahwa setiap observasi ke- i berisi sepasang p prediktor $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$, $i = 1, 2, \dots, n$, dan terkait dengan $y_i = \{-1, 1\}$. Maka didapatkan himpunan

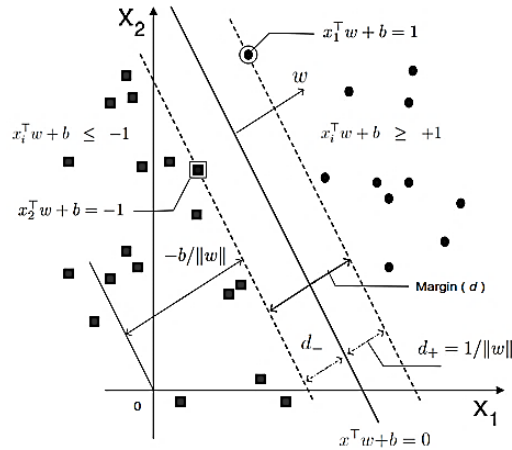
$$\mathbf{D}_n = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \in X \times \{-1, 1\},$$

Domain X adalah dataset tidak kosong \mathbf{x}_i , dan y_i adalah label bertanda negatif atau positif. Indeks $i = 1, 2, \dots, n$ digunakan saat melatih data. Jika x_i adalah anggota kelas (+1) maka x_i diberi label $y_i = +1$ (bertanda positif) dan jika tidak maka diberi label $y_i = -1$ (bertanda negatif) sehingga data yang diberikan berupa pasangan $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ merupakan himpunan data dari dua kelas yang akan diklasifikasikan dengan SVM (Gunn, 1998).

Konsep utama untuk mendefinisikan sebuah pengelompokan linier ialah perkalian titik, antara dua vektor yang ditetapkan sebagai $\mathbf{x}^T \mathbf{w} = \sum_{i=1}^n \mathbf{x}_i \mathbf{w}_i$. Pada Gambar 2.2, dapat dilihat bahwa berbagai alternatif batas pemisah yang dapat memisahkan semua dataset sesuai dengan kelasnya namun batas pemisah terbaik tidak hanya dapat memisahkan data tetapi juga memiliki margin paling besar (Borges, 1998). Fungsi klasifikasi $\mathbf{x}^T \mathbf{w} + b$ pada Gambar 2.2 berada dalam sebuah keluarga fungsi klasifikasi F yang terbentuk, yaitu

$$F = \{\mathbf{x}^T \mathbf{w} + b, \mathbf{w} \in \mathbb{R}^p, b \in \mathbb{R}\}$$

di mana \mathbf{w} merupakan vektor pembobot dan b adalah *bias*.



Gambar 2.2 Bidang pemisah terbaik dengan margin (m) terbesar (Haerdle, Prastyo, & Hafner, 2014)

Diberikan batas pemisah (*separating hyperplane*)

$$f(x) = \mathbf{x}^T \mathbf{w} + b = 0 \quad (2.3)$$

yang membagi ruang (*space*) menjadi dua wilayah seperti pada Gambar 2.2. Bentuk pada $f(x)$ adalah semua titik yang tegak lurus terhadap \mathbf{w} dan melalui konstanta b . Bentuk pada $f(x)$ merupakan sebuah garis dalam dua dimensi, sebuah bidang pada tiga dimensi, dan secara umum berupa *hyperplane* pada dimensi yang lebih tinggi. *Hyperplane* dikatakan linier jika merupakan fungsi linier dalam input x_i sehingga disebut *linier classifier*. Penentuan vektor pendukung membutuhkan *margin* yang maksimal. *Margin* adalah koridor di antara batas pemisah. Data yang berada pada *margin (m)* disebut dengan *support vector*. Sebab pemisahan bisa dilakukan tanpa *error*, setiap observasi $i = 1, 2, \dots, n$ harus memenuhi kendala pemisah untuk kedua kelas adalah sebagai berikut:

$$\mathbf{x}^T \mathbf{w} + b \geq 1 \text{ untuk } y_i = +1 \quad (2.4.a)$$

$$\mathbf{x}^T \mathbf{w} + b \leq -1 \text{ untuk } y_i = -1, \quad (2.4.b)$$

Dimana adalah w vektor bobot (*weight vector*) yang berukuran $(p \times 1)$, b adalah posisi bidang relatif terhadap pusat koordinat atau lebih dikenal dengan bias yang bernilai skalar.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad \mathbf{X}^T = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{bmatrix}$$

$$\mathbf{x}_i^T = [x_{i1} \quad x_{i2} \quad \dots \quad x_{ip}] \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Pada Gambar 2.2 menunjukkan $\frac{|b|}{\|\mathbf{w}\|}$ adalah jarak batas pemisah yang tegak lurus dari titik pusat koordinat dan $\|\mathbf{w}\|$ adalah panjang vektor Eulidean (*Eulidean's norm*) yang merupakan jarak dari \mathbf{w} . Panjang vektor *euclidean* dari \mathbf{w} adalah $\sqrt{\mathbf{w}^T \mathbf{w}} = \sqrt{w_1^2 + w_2^2 + \dots + w_p^2}$. Batas pertama membatasi kelas (+1) sedangkan pembatas kedua membatasi kelas (-1).

Batas pertama $\mathbf{x}_i^T \mathbf{w} + b = +1$ mempunyai bobot \mathbf{w} dan jarak tegak lurus dari titik asal sebesar $\frac{|1-b|}{\|\mathbf{w}\|}$, sedangkan batas kedua $\mathbf{x}_i^T \mathbf{w} + b = -1$ mempunyai bobot \mathbf{w} dan jarak tegak lurus dari titik asal sebesar $\frac{|-1-b|}{\|\mathbf{w}\|}$. Nilai maksimum *margin* antara pembatas (berdasarkan rumus jarak garis ketitik pusat) adalah

$$\frac{1-b-(-1-b)}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|} \quad (2.5)$$

Margin pada *hyperplane* $f(x)$ dengan memasukkan data set \mathbf{D}_n mengikuti Persamaan (2.6).

$$d_D f = \frac{1}{2} \mathbf{w}^T (x_+ - x_-) \quad (2.6)$$

dimana unit vektor \mathbf{w} adalah arah pada \mathbf{w} . Ini diasumsikan bahwa x_+ dan x_- memiliki jarak yang sama dari bidang pembatas (*separating hyperplane*). Mengoptimalkan *hyperplane* yaitu dengan cara mengoptimalkan Persamaan (2.3) atau dengan meminimumkan fungsi pada persamaan (2.7)

$$\min \frac{1}{2} \|\mathbf{w}\|^2 \quad (2.7)$$

dengan cara menggabungkan kedua pembatas pada persamaan (2.4.a) dan (2.4.b) yang menjadi pertidaksamaan sebagai berikut:

$$y_i (\mathbf{x}_i^T \mathbf{w} + b) - 1 \geq 0, \quad i = 1, 2, \dots, n \quad (2.8)$$

Persoalan optimasi untuk Persamaan (2.3) lebih mudah diselesaikan jika diubah ke dalam formula *Lagrange*. Dengan demikian permasalahan optimasi dengan pembatas dapat dirumuskan menjadi:

$$\min_{\mathbf{w}, b} L_p(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i \{y_i(\mathbf{x}_i^T \mathbf{w} + b) - 1\} \quad (2.9)$$

dengan kendala pembatas $\alpha_i \geq 0$ (nilai dari koefisien *Lagrange*). Penaksir \mathbf{w} dan b dengan meminimumkan L_p terhadap \mathbf{w} dan b dan disamadengankan dengan nol yaitu $\frac{\partial L_p(\mathbf{w}, b, \alpha)}{\partial \mathbf{w}} = 0$ dan $\frac{\partial L_p(\mathbf{w}, b, \alpha)}{\partial b} = 0$, sehingga diperoleh Persamaan (2.10) dan (2.11)

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \quad (2.10)$$

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad (2.11)$$

Vector \mathbf{w} seringkali bernilai tak terhingga, tetapi nilai α_i terhingga. Maka, formula *Lagrange* L_p (*primal problem*) diubah ke dalam L_D (*Dual Problem*). Dengan mensubstitusikan Persamaan (2.7) ke Persamaan (2.8) diperoleh L_D yang ditunjukkan pada persamaan (2.12) :

$$L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \quad (2.12)$$

Jadi persoalan pencarian bidang pemisah terbaik dapat dirumuskan pada persamaan (2.13).

$$\max_a L_D \quad (2.13)$$

Nilai α_i yang diperoleh, nantinya akan digunakan untuk mencari nilai \mathbf{w} . Jika nilai $\alpha_i \geq 0$ maka pengamatan x_i merupakan sebuah *support vector*. Setelah menyelesaikan *Dual Problem*, maka suatu pengamatan baru \mathbf{x}_{new} dapat diklasifikasikan menggunakan ukuran klasifikasi sebagai persamaan (2.14).

$$\hat{f}(\mathbf{x}_{new}) = \text{sign}(\mathbf{x}_{new}^T \hat{\mathbf{w}} + \hat{b}), \quad (2.14)$$

dimana

$$\hat{\mathbf{w}} = \sum_{i=1}^n \alpha_i y_i \mathbf{X}_i,$$

$$\hat{b} = \frac{1}{n_{sv}} \left(\sum_{i=1}^{n_{sv}} \frac{1}{y_i} - (\mathbf{x}_{new}^T \hat{\mathbf{w}}) \right),$$

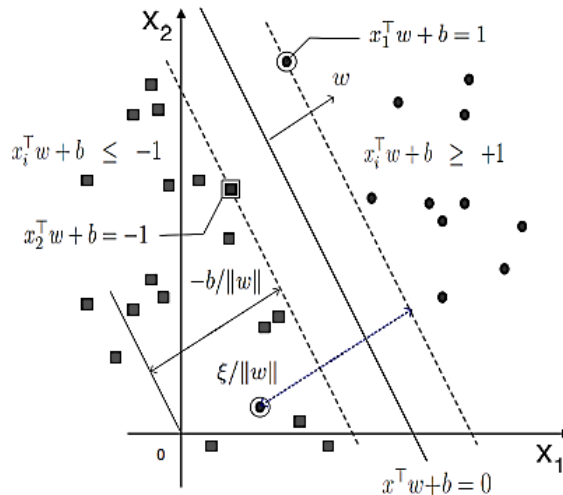
\mathbf{X}_i adalah *support vector*, \mathbf{x}_{new} merupakan data yang akan diklasifikasikan, α_i adalah *Lagrange Multiplier*, b adalah bias, dan n_{sv} adalah jumlah *support vector*.

2.5.2 SVM Linear Nonseparable

Haerdle, dkk (2014) menyatakan pada kasus linier *nonseparable* yaitu mengklasifikasikan data linier yang tidak dapat dipisahkan, maka kendala pemisah pada persamaan (2.4.a) dan (2.4.b) harus diubah secara linier dengan penambahan variabel *slack* ξ_i , yang menunjukkan penalti terhadap ketelitian pemisahan yang memungkinkan suatu titik berada di antara *margin error* ($0 \leq \xi_i \leq 1, \forall_i$) atau $\xi_i > 1$. Penalti ketelitian kemungkinan terjadinya keadaan suatu titik berada di antara *margin error* disebut *missclassified*. Sehingga x_i diklasifikasikan menjadi :

$$x_i^T \mathbf{w} + b \geq 1 - \xi_i \text{ untuk } y_i = +1 \quad (2.15.a)$$

$$x_i^T \mathbf{w} + b \geq -(1 - \xi_i) \text{ untuk } y_i = -1 \quad (2.15.b)$$



Gambar 2.3 Bidang pemisah terbaik dengan margin terbesar linier *non separable* (Haerdle, Prastyo, & Hafner, 2014)

Bidang pemisah terbaik dengan *margin* terbesar linier *nonseparable*, diilustrasikan pada Gambar 2.3. Pencarian bidang pemisah terbaik dengan penambahan variabel ξ_i sering juga disebut dengan *soft margin hyperplane*. Formula pencarian bidang pemisah terbaik berubah menjadi :

$$\min_{\mathbf{w}, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C[\sum_{i=1}^n \xi_i] \quad (2.16)$$

Persamaan (2.15.a) dan (2.15.b) dapat diubah dalam bentuk Persamaan (2.17) :

$$y_i(x_i^T \mathbf{w} + b) \geq 1 - \xi_i \quad (2.17)$$

dengan $\xi_i \geq 0, C > 0$ dimana C adalah parameter yang menentukan besar penalti akibat kesalahan dalam klasifikasi (*misclassification*) dari data *training* dan

nilainya ditentukan oleh peneliti. Bentuk persamaan (2.16) memenuhi prinsip *Structural Risk Minimization* (SRM). Prinsip SRM ialah meminimumkan $\frac{1}{2} \|\mathbf{w}\|^2$ ekuivalen dengan meminimumkan dimensi VC (Vapnik-Chervonenkis). Nilai dari dimensi VC ini akan menentukan besarnya nilai kesalahan hipotesis pada data *testing* sedangkan meminimumkan $C[\sum_{i=1}^n \xi_i]$ ekuivalen dengan meminimumkan *error* pada data *training*. Fungsi *Langrange* untuk *primal problem* adalah

$$L_p(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i (\mathbf{x}_i^T \mathbf{w} + b) - 1 + \xi_i] - \sum_{i=1}^n \mu_i \xi_i \quad (2.18)$$

dimana $\alpha_i \geq 0$ dan $\mu_i \geq 0$ adalah *Lagrange Multiplier*. Kondisi KKT (*Karush-Kuhn-Tucker*) untuk *primal problem* adalah

$$\frac{\partial L_p(\mathbf{w}, b, \alpha)}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i = 0 \leftrightarrow \mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \quad (2.19)$$

$$\frac{\partial L_p(\mathbf{w}, b, \alpha)}{\partial b} = - \sum_{i=1}^n \alpha_i y_i = 0 \leftrightarrow \sum_{i=1}^n \alpha_i y_i = 0 \quad (2.20)$$

$$\frac{\partial L_p(\mathbf{w}, b, \alpha)}{\partial \alpha} = C - \alpha_i - \mu_i \leftrightarrow C = \alpha_i + \mu_i \quad (2.21)$$

dengan kendala:

$$\xi_i \geq 0,$$

$$\alpha_i \geq 0,$$

$$\mu_i \geq 0,$$

$$\alpha_i \{y_i (\mathbf{x}_i^T \mathbf{w} + b) - 1 + \xi_i\} = 0, \text{ dan}$$

$$\mu_i \xi_i = 0$$

juga men-sibstusikan nilai $\hat{\mathbf{w}} = \sum_{i=1}^n \hat{\alpha}_i y_i \mathbf{x}_i$ ke dalam *primal problem* menjadi persamaan *dual problem* sebagai berikut :

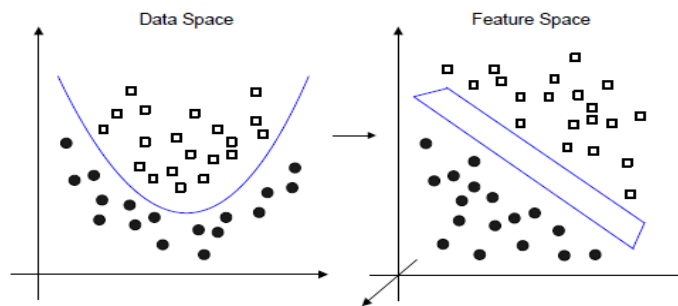
$$\max_{\alpha} L_D = \max \left(\sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \right) \quad (2.22)$$

$$\text{dengan } 0 \leq \alpha_i \leq C, \text{ dan } \sum_{i=1}^n \alpha_i y_i = 0$$

Sampel \mathbf{x}_i untuk $\alpha_i > 0$ (*support vector*) yaitu titik yang berada di luar *margin* atau dalam *margin* ketika *soft margin* digunakan. *Support vector* sering menyebar dan level penyebaran berada pada batas atas (*upper bound*) untuk tingkat misklasifikasi (Scholkopf & Smola, 2002).

2.5.3 SVM *Nonlinear Separable*

Menurut Haerdle, dkk (2014), pada kenyataan tidak semua data bersifat linier sehingga sulit untuk mencari bidang pemisah secara linier. Untuk mengklasifikasikan data yang tidak linier dapat diatasi dengan SVM Non-linier. Diberikan beberapa titik baru $x \in X$ dan ingin memprediksi hubungan $y \in Y = \{-1,1\}$. Permasalahan non-linier dapat diselesaikan dengan mentransformasikan data ke dalam ruang yang berdimensi lebih tinggi, sehingga dapat dipisahkan secara linier pada *feature space* yang baru.



Gambar 2.4 Mapping dari Dua Dimensi *Data Space* (Kiri) ke Tiga Dimensi *Feature Space* (Kanan) (Haerdle, Prastyo, & Hafner, 2014)

Klasifikasi non-linier yang ditunjukkan pada Gambar 2.4, data dengan sebuah struktur non-linier fungsi $\varphi: \mathbb{R}^p \rightarrow H$ ke dalam *dimensional space* tinggi H dimana pengukuran klasifikasi bersifat linier. Semua *vector training* x_i dalam persamaan (2.22) berupa *scalar product* dengan bentuk $\mathbf{X}_i^T \mathbf{X}_j$. Pada SVM non-linier, *scalar product* ditransformasikan ke $\varphi(\mathbf{X}_i)^T \varphi(\mathbf{X}_j)$. Fungsi transformasi pada SVM ialah menggunakan “*Kernel Trick*” (Scholkopf & Smola, 2002). *Kernel Trick* adalah menghitung *scalar product* dalam bentuk sebuah fungsi kernel. Proyeksi $\varphi: \mathbb{R}^p \rightarrow H$ memastikan bahwa *inner product* $\varphi(\mathbf{X}_i)^T \varphi(\mathbf{X}_j)$ dipresentasikan oleh fungsi kernel

$$K(\mathbf{X}_i \mathbf{X}_j) = \varphi(\mathbf{X}_i)^T \varphi(\mathbf{X}_j) \quad (2.23)$$

Fungsi kernel K pada persamaan (2.23) dapat digunakan tanpa perlu mengetahui fungsi transformasi φ secara eksplisit.

Diberikan sebuah kernel K dan beberapa himpunan data $x_1, x_2, \dots, x_n \in X$,

maka matrik $K = \left(K(\mathbf{x}_i \mathbf{x}_j) \right)_{ij}$ dengan ukuran $n \times n$ disebut *Gram matrix* untuk data x_1, x_2, \dots, x_n . Sebuah syarat cukup dan perlu untuk matrik simetri K , dengan $K_{ij} = K(\mathbf{x}_i \mathbf{x}_j) = K(\mathbf{x}_j \mathbf{x}_i) = K_{ji}$ untuk K definit positif disebut “*Mercer’s Theorem*” (Mercer & Li, 1909).

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j K(\mathbf{x}_i \mathbf{x}_j) \geq 0 \quad (2.24)$$

Contoh sederhana pada sebuah *kernel trick* yang menunjukkan bahwa kernel dapat dihitung tanpa perhitungan fungsi *mapping* φ secara eksplisit adalah fungsi pemetaan :

$$\varphi(x_1, x_2) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)^T \quad (2.25)$$

Sehingga menjadi

$$\varphi^T(x) \varphi(x) = \mathbf{w}_1 x_1^2 + \sqrt{2} \mathbf{w}_2 x_1 x_2 + \mathbf{w}_3 x_2^2 \quad (2.26)$$

dengan dimensi pada *feature space* berupa kuadratik, padahal dimensi asalnya adalah linier. Metode kernel menghindari pembelajaran secara eksplisit *mapping* data ke dalam *feature space* dimensi tinggi, seperti pada contoh berikut:

$$\begin{aligned} f(x) &= \mathbf{w}^T \mathbf{x} + b \\ &= \sum_{i=1}^n \alpha_i \mathbf{x}_i^T \mathbf{x} + b \\ &= \sum_{i=1}^n \alpha_i \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}) + b \text{ dalam } \textit{feature space } \mathcal{F} \\ &= \sum_{i=1}^n \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b \end{aligned}$$

Hubungan kernel dengan fungsi *mapping* adalah:

$$\begin{aligned} \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}) &= (x_{i1}^2, \sqrt{2}x_{i1}x_{i2}x_{i2}^2)(x_{i1}^2, \sqrt{2}x_{i1}x_{i2}x_{i2}^2)^T \\ &= x_{i1}^2 x_1^2 + 2x_{i1}x_{i2}x_1x_2 + x_{i2}^2 x_2^2 \\ &= (\mathbf{x}_i^T \mathbf{x})^2 \\ &= K(\mathbf{x}_i, \mathbf{x}_j) \end{aligned}$$

Sedangkan, untuk memperoleh fungsi klasifikasi non-linier dalam data *space*, bentuk secara umumnya dipeoleh dari penerapan *kernel trick* ke Persamaan (2.27):

$$L_D(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_j y_i y_j K(\mathbf{x}_i \mathbf{x}_j) \quad (2.27)$$

yaitu memaksimumkan Persamaan (2.28)

$$\max_{\alpha} L_D = \max \left(\sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i \mathbf{x}_j) \right) \quad (2.28)$$

dengan $\sum_{i=1}^n \alpha_i y_i = 0$, $0 \leq \alpha_i \leq C$, $i = 1, 2, \dots, n$.

Data jenis non-linier dapat diselesaikan dengan metode kernel. Fungsi Kernel akan memetakan data ke dalam *space* dengan dimensi yang lebih tinggi sehingga mampu untuk melakukan separasi data. Konsep dengan menggunakan fungsi kernel ini sangat baik. Fungsi kernel mendukung SVM untuk memisahkan data dengan batasan yang sangat rumit. Fungsi hasil training yang dihasilkan yakni seperti persamaan (2.29).

$$f(x_k) = \sum_{i=1}^n \alpha_i y_i K(\mathbf{x}_i; \mathbf{x}_j) + b \quad (2.29)$$

Fungsi kernel yang umum digunakan pada metode SVM adalah :

1. Kernel Linier

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j \quad (2.30)$$

2. Kernel Polynomial

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\delta \mathbf{x}_i^T \mathbf{x}_j + r)^P, \delta > 0 \quad (2.31)$$

3. Kernel Radial basis function (RBF)

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \quad (2.32)$$

Dengan $\|\mathbf{x}_i - \mathbf{x}_j\|^2$ dikenal sebagai *squared Euclidean distance* diantara dua vektor variabel fitur dan σ adalah variabel bebas. Definisi yang setara tetapi lebih sederhana, melibatkan parameter $\gamma = \frac{1}{2\sigma^2}$.

4. Kernel sigmoid

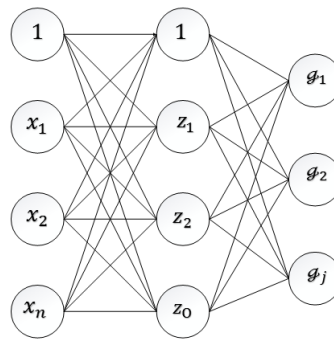
$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\mathbf{x}_i^T \mathbf{x}_j + r) \quad (2.33)$$

Pemilihan fungsi kernel yang tepat merupakan hal yang sangat penting karena akan menentukan *feature space* dimana fungsi *classifier* akan dicari. Sepanjang fungsi kernelnya sesuai (cocok), SVM akan beroperasi secara benar meskipun tidak tahu pemetaan yang digunakan (Santosa, 2007; Robandi, 2008). Banyak fungsi kernel yang dapat digunakan, tetapi hanya sedikit fungsi yang dapat bekerja dengan baik dalam aplikasi yang sangat luas. Fungsi yang paling sering digunakan dan sangat direkomendasikan adalah *Radial Basis Function* (RBF). Menurut Hsu (2004), fungsi kernel yang direkomendasikan untuk diuji pertama kali adalah fungsi kernel RBF karena memiliki performansi yang sama dengan kernel linier pada parameter tertentu, memiliki perilaku seperti fungsi kernel sigmoid dengan parameter tertentu dan rentang nilai yang kecil [0,1].

2.6 Neural Network

Model *Neural Network* dengan jaringan *backpropagation* ialah salah satu metode *neural network* dengan ciri khas jaringan layar jamak (*multi layer network*). Ciri khasnya yaitu memiliki tiga jenis *layer*, yakni *input layer*, *hidden layer*, dan *output layer*. Jaringan seperti ini dapat menyelesaikan permasalahan yang lebih kompleks daripada *neural network* layar tunggal, tetapi prosesnya sering memakan waktu yang lebih banyak. Model *backpropagation* menggunakan cara pelatihan *supervised learning*. Menurut cara pelatihan ini, setiap pola yang diberikan ke dalam *artificial neural network* telah diketahui *output*-nya. Selisih antara pola aktual (*output* yang dihasilkan) dengan pola *output* yang dikehendaki (*output target*) yang disebut *error*, digunakan untuk mengoreksi bobot tiruan. Hal itu dilakukan sehingga *artificial neural network* mampu menghasilkan *output* sedekat mungkin dengan pola target yang telah diketahui. Contoh algoritma *artificial neural network* yang menggunakan cara pelatihan *supervised learning* selain metode *backpropagation* adalah: Hebbian, Perceptron, ADALINE, Boltzman, dan Hopfield.

Metode *Backpropagation* terdiri dari banyak neuron dengan kemampuan non-linier *mapping*. Metode *Backpropagation* merupakan salah satu metode yang sederhana dan umum digunakan untuk *supervised training* pada *multilayer artificial neural networks*. Metode ini bekerja dengan memperkirakan hubungan non-linier antara *input* dan *output* dengan menyesuaikan nilai sesuai dengan nilai minimum dari *error function* sehingga memungkinkan jaringan untuk berpusat pada keadaan stabil dan memberikan *output* yang sesuai ketika menerima masukan yang tidak termasuk ke dalam pola data *training*. Secara umum, jaringan *backpropagation* memiliki dua tahap, *training* dan *testing*. Pada tahap *training*, jaringan melakukan proses dengan pola input dan klasifikasi yang benar. Pada tahap *testing*, input yang digunakan adalah *input* baru yang tidak termasuk ke dalam pola *input training*, sehingga akan diketahui hasil klasifikasi berdasarkan tahap *training*. Arsitektur *backpropagation* diilustrasikan seperti pada Gambar 2.5.



Gambar 2.5 Arsitektur *Backpropagation*

Disiplin kerja jaringan *backpropagation* terbagi menjadi dua tahapan yaitu *feedforward propagation* dan *backward propagation*. Misalnya terdapat suatu set pelatihan berupa vektor *input* dan vektor *output* target. Keluaran yang diperoleh berupa vektor *output* aktual. Setelah itu kemudian dilakukan perbandingan antara *output* aktual dengan *output* target dengan cara mencari selisih diantara kedua *output* tersebut. Selisih yang didapatkan adalah *error* yang dijadikan sebagai dasar dalam melakukan perubahan dari setiap bobot dengan cara mempropagasi kembali. Perubahan bobot dapat mempengaruhi *error*. Siklus perubahan bobot dilakukan pada setiap set pelatihan hingga bila mencapai jumlah bobot yang diinginkan atau hingga nilai ambang yang ditetapkan terlampaui.

Tahapan pengerjaan algoritma *backpropagation* secara detail (Agustin & Prahasto, 2012) ialah sebagai berikut:

Langkah 1 : Inisiasi bobot-bobot, konstanta laju pelatihan (Δ), toleransi *error* atau nilai bobot atau menentukan maksimal *epoch*.

Langkah 2 : Selama kondisi berhenti belum sesuai, kerjakan langkah ke-2 sampai langkah 10.

Langkah 3 : Untuk setiap pasangan pola pelatihan, lakukan langkah ke-4 sampai langkah ke-9.

Langkah 4: (Tahap I : *feedforward*) Tiap unit *input* ($x_i, i = 1, 2, \dots, I$) menerima sinyal dan meneruskannya ke unit *hidden layer*.

Langkah 5: Masing-masing unit di *hidden layer* ($z_o, o = 1, 2, \dots, O$) dikalikan dengan bobotnya dan dijumlahkan serta ditambahkan dengan biasanya.

Bobot sinyal *input*:

$$z_{in_o} = v_{0o} + \sum_i^I x_i v_{io} \quad (2.34)$$

Dengan menerapkan fungsi aktivasi *sigmoid*:

$$f(z) = \frac{1}{1+e^{-z_{in_o}}} \quad (2.35)$$

Langkah 6 : Masing-masing unit *output* ($g_j, j = 1, 2, \dots, J$) dikalikan dengan bobot (2.34), dijumlahkan bobot sinyal *input* dengan menerapkan fungsi aktivasi input, serta ditambahkan dengan biasnya. Lalu mengirimkan isyarat ini ke semua unit diteruskan menuju *output*.

Bobot dari *hidden layer*:

$$g_{in_j} = u_{0j} + \sum_k^K z_{ok} u_{kj} \quad (2.36)$$

Dengan menerapkan fungsi aktivasi *sigmoid*:

$$f(g) = \frac{1}{1+e^{-g_{in_j}}} \quad (2.37)$$

Langkah 7 : (Tahap II : *backward propagation*) Masing-masing unit *output* ($g_j, j = 1, 2, \dots, J$) menerima pola pelatihan dan kemudian informasi *error* ϵ_j lapisan *output* dihitung. Selisih tersebut dikirim ke lapisan *hidden layer* dan digunakan untuk menghitung besarnya koreksi bobot bias antara lapisan *hidden layer* dengan lapisan *output*. Perhitungan error:

$$\epsilon_j = (\vartheta_j - g_j) f'(g), \text{ dengan } \vartheta_j = u_{j=_{u^{layer+1}} \epsilon^{layer+1}}$$

Perhitungan koreksi bobot dan bias berdasarkan *error* di setiap *hidden layer* :

$$\Delta u_{oj} = \epsilon_j z_o \quad (2.38)$$

$$\Delta u_{0j} = \epsilon_j \quad (2.39)$$

Langkah 8: Tiap unit di *hidden layer* ($z_o, o = 1, 2, \dots, O$) menjumlahkan error masukannya (dari unit-unit yang berada pada lapisan atasnya). Kemudian dihitung *error* dan koreksi bobot biasnya.

$$\epsilon_k = \epsilon_{in_o} f'(z) \quad (2.40)$$

$$\Delta u_{io} = \epsilon_o x_i \quad (2.41)$$

Langkah 9 : Tiap unit *output* ($\varphi_j, j = 1, 2, \dots, J$) memperbaharui bobot dan bias,

$$u_{oj}(\text{baru}) = u_{oj}(\text{lama}) + \Delta u_{io} \quad (2.42)$$

dan tiap unit *hidden layer* ($z_o, o = 1, 2, \dots, O$) memperbaharui bobot dan biasnya.

$$v_{io}(\text{baru}) = v_{io}(\text{lama}) + \Delta v_{io} \quad (2.43)$$

Langkah 10 : Uji syarat berhenti

Pada Widhianingsih (2016), disebutkan bahwa algoritma *backpropagation* dapat melakukan pelatihan hingga didapatkan hasil yang paling optimum. Hasil tersebut didapatkan dengan cara meminimalisasi besar *error* yang ada. Berdasarkan algoritma ini, *error* tersebut diminimalisasi dengan cara memodifikasi pembobot yang digunakan di dalam perhitungan. Ketika ditemukan bobot baru, proses pelatihan akan dilakukan kembali. Proses ini akan dilakukan secara berulang-ulang dan akan berhenti pada saat didapatkan *error* yang telah memenuhi uji syarat berhenti.

Proses meminimalisasi besar *error* dilakukan dengan meminimalkan fungsi resiko empiris atau fungsi biaya (*cost*). Fungsi biaya dinyatakan seperti pada Persamaan (2.44)

$$B = \frac{1}{2N} \sum_{n=1}^N \sum_{j=1}^J \left(\varphi_j(n) - \hat{\varphi}_j(n) \right)^2 \quad (2.44)$$

Dengan φ_j merupakan target dan $\hat{\varphi}_j$ adalah prediksi kategori, serta N adalah jumlah data training. Berdasarkan Persamaan (2.44), *loss function* dinyatakan dengan $\left(\varphi_j(n) - \hat{\varphi}_j(n) \right)^2$. Tujuan menggunakan *loss function* adalah untuk mengetahui tingkat kesalahan yang ditimbulkan dari model yang didapatkan.

2.7 Cross Validation

Metode *Cross Validation* adalah metode statistik yang digunakan untuk mengevaluasi dan membandingkan algoritma yang dipelajari dengan cara membagi data menjadi dua bagian: satu digunakan untuk mempelajari atau melatih model dan satunya digunakan untuk memvalidasi model. Bagian untuk mempelajari algoritma dan melatih model ialah *training set* sedangkan bagian untuk memvalidasi model dan algoritma yang telah dipelajari disebut *testing set*. Dalam

tipikal *cross-validation*, *training set* dan *testing set* harus saling silang berturut-turut hingga setiap titik data memiliki kesempatan untuk tervalidasi.

Tujuan dari *cross-validation* ada dua. Pertama, mengukur kinerja model pembelajaran dari data yang tersedia menggunakan suatu algoritma. Singkatnya, untuk mengukur generalisasi sebuah algoritma. Kedua, untuk membandingkan kinerja dua atau lebih kinerja yang berbeda dan mengetahui algoritma terbaik untuk data yang ada, atau alternatif lain untuk membandingkan kinerja dua atau lebih varian model parameter.

Bentuk lain dari *cross-validation* yaitu kasus spesial dari *cross-validation* atau menyangkut putaran berulang *cross-validation*. Pada *cross-validation*, pertama-tama data dipartisi sebanyak q bagian sama rata atau hampir sama rata. Selanjutnya, iterasi ke- q *training set* dan *testing set* dilakukan sedemikian rupa sehingga dalam setiap iterasi, setiap lipatan (*fold*) data yang berbeda akan diadakan validasi (*testing*), sedangkan sisa lipatannya akan digunakan untuk pembelajaran (*training*).

Data biasanya berlapis stratifikasi lebih dulu sebelum dibagi menjadi q bagian. Lapisan stratifikasi adalah proses penataan ulang data untuk memastikan setiap lipatan merupakan perwakilan yang baik. Misalnya dalam masalah klasifikasi biner dimana masing-masing kelas terdiri dari 50% data. Sebaiknya data diatur sedemikian rupa sehingga di setiap lipatan, masing-masing kelas terdiri dari setengah kasus.

Dalam setiap iterasi, satu atau lebih algoritma pembelajaran menggunakan $q - 1$ lipatan data untuk mempelajari satu atau lebih model, dan selanjutnya model yang telah dipelajari akan diminta untuk membuat prediksi tentang data dalam lipatan validasi. Kinerja setiap pembelajaran algoritma pada setiap lipatan dapat dilacak dengan menggunakan beberapa metrik kinerja yang telah ditentukan seperti akurasi. Berdasarkan penyelesaian, metrik kinerja q sampel akan tersedia untuk setiap algoritma.

2.8 Teknik Evaluasi Kinerja Model Klasifikasi

Menurut Sarkar (2016), membentuk model memang bagian yang penting didalam seluruh analisis, tetapi lebih penting lagi ialah mengetahui seberapa baik

model-model tersebut dipersembahkan. Tampilan model klasifikasi biasanya berdasarkan seberapa baik model tersebut memprediksi hasil untuk titik data yang baru. Biasanya tampilan ini diukur terhadap sebuah tes atau dataset yang berisi titik-titik data yang tidak digunakan untuk mempengaruhi atau melatih *classifier*. Tes dataset seperti ini biasanya memiliki beberapa observasi dan label yang sesuai.

Machine Learning (ML) membagi klasifikasi ke dalam bentuk biner, multi-kelas, multi-label, dan hierarki. *Supervised* ML memungkinkan akses ke label data selama tahap *training* dan *testing*. Saat menginput data x_1, x_2, \dots, x_n , label kategorik harus ditempatkan ke kelas yang telah ditentukan. Untuk kasus biner, inputnya harus diklasifikasikan menjadi dua kelas yang tidak tumpang tindih. Kategori yang ditentukan bisa objektif (tidak tergantung pada evaluasi manual), atau subjektif (bergantung pada evaluasi manual, misalnya ulasan positif atau negatif). Kelas bisa didefinisikan dengan baik (contoh: label pungut suara), ambigu (contoh: ulasan label pendapat), atau keduanya (contoh: teks medis) (Sokolova & Lapalme, 2009).

Sebagai sebuah ilmu, ML meminjam langkah-langkah dari bermacam-macam disiplin yang secara tradisional bergantung pada bukti empiris. Dalam beberapa hal, klasifikasi teks meminjam dari *Information Extraction* (IE) yang mengawali penggunaan ML dalam pemrosesan dan pemahaman teks otomatis, misalnya, analisis otomatis dan pembuatan teks sinonim. Metrik IE dan *Information Retrieval* (IR) dalam evaluasi algoritma ML ialah contoh peminjaman tersebut. Metrik yang dimaksud adalah metrik evaluasi umum yang digunakan dalam klasifikasi teks (Sokolova & Lapalme, 2009). Beberapa perhitungan yang menentukan performa dari prediksi model dalam klasifikasi teks ialah: akurasi, presisi, sensitivitas, dan skor F1. Ketepatan klasifikasi dapat dievaluasi dengan menghitung jumlah sampel kelas yang diakui dengan benar (*True Positive* atau TP), jumlah sampel yang diakui dengan benar namun tidak digolongkan dalam kelas (*True Negative* atau TN), sampel yang salah penetapan kelas (*False Positive* atau FP) atau yang tidak dikenali sebagai sampel kelas (*False Negative* atau FN). Keempat perhitungan ini merupakan *confusion matrix* yang ditunjukkan pada Gambar 2.6 untuk kasus klasifikasi biner.

		prediction outcome		P'	N'
		p	n		
actual value	p'	True Positive	False Negative		
	n'	False Positive	True Negative		
		P	N		

Gambar 2.6 *Confusion Matrix* dua kelas klasifikasi

Menurut Sarkar (2016), Akurasi didefinisikan sebagai akurasi keseluruhan atau proporsi prediksi benar model, bisa diformulasikan sebagai (2.45).

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \quad (2.45)$$

Presisi atau *specificity* didefinisikan sebagai jumlah prediksi yang benar benarnya atau relevan dari semua prediksi berdasarkan kelas positif. Hal ini juga diketahui sebagai nilai prediktif positif dan bisa digambarkan dengan formulasi (2.46).

$$\text{Presisi} = \frac{TN}{TN+FP} \quad (2.46)$$

Recall didefinisikan sebagai jumlah data kelas positif yang diprediksi dengan benar. Istilah *recall* juga bisa dikenal sebagai *hit rate*, *coverage*, atau *sensitivity* dan bisa digambarkan dengan formulasi (2.47).

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2.47)$$

Skor F1 adalah pengukuran akurasi yang lain yang dihitung dengan cara mengambil rata-rata harmonik dari presisi dan sensitivitas dan bisa dijelaskan dengan formula (2.48).

$$\text{Skor F1} = 2 / \left(\frac{1}{\text{presisi}} + \frac{1}{\text{sensitivitas}} \right) \quad (2.48)$$

2.9 Pembentukan Kluster dengan Metode K-Means

K-Means merupakan salah satu algoritma pembentukan kluster yang bertujuan untuk membagi data ke dalam beberapa kelompok. Algoritma ini menerima masukan berupa data tanpa label kelas. Hal inilah yang membedakan pembentukan kluster yang *unsupervised learning* dan klasifikasi yang *supervised learning* dimana label kelas kelompoknya telah diketahui sebelumnya.

Algoritma kluster K-Means yang digunakan pada penelitian ini adalah Algoritma Hartigan dan Wong (Morisette & Chartier, 2013). Algoritma tersebut mencari ruang data dengan optimal lokal dalam jumlah kesalahan kluster (*locally optimal within-kluster sum of square of errors*). Berarti, optimal lokal *within-kluster* dapat memberikan kasus ke subruang lain, bahkan jika saat ini optimal lokal *within-kluster* milik subruang dari *centroid* terdekat akan meminimalkan jumlah total kluster dalam kuadrat (*the total within-kluster sum of square*).

Iterasinya adalah sebagai berikut. Jika *centroid* telah diperbarui pada langkah terakhir, untuk setiap titik data yang disertakan, jumlah kuadrat di dalam kluster untuk setiap titik data apabila termasuk kelompok lain yang terhitung. Jika salah satu dari jumlah kuadrat (SSE2) lebih kecil dari yang sekarang (SSE1) untuk semua $i \neq 1$ seperti pada persamaan (2.49), kasus ditugaskan ke kluster baru ini.

$$SSE2 = \frac{N_i \sum_j \|x_{ij} - c_i\|^2}{N_i - 1} < SSE1 = \frac{N_1 \sum_j \|x_{1j} - c_1\|^2}{N_1 - 1} \quad (2.49)$$

Iterasi terus berlanjut hingga tidak ada perubahan gugus kasus, yang berarti sampai rentang waktu tertentu akan membuat kluster lebih bervariasi secara internal atau lebih mirip secara eksternal. Berikut dipaparkan langkah-langkah algoritma Hartigan dan Wong:

- a. Pilih jumlah kluster.
- b. Pilih metrik yang akan digunakan.
- c. Pilih metode untuk memilih *centroid* awal.
- d. Tentukan *centroid* awal.
- e. Tetapkan kasus ke *centroid* terdekat.
- f. Hitung *centroid*.
- g. Kluster $j \leq$ jumlah kluster, Jika *centroid* j diperbaharui diiterasi terakhir, maka
 - i. Hitung SSE *within* kluster
 - ii. Untuk kasus $i \leq$ *within* kluster
 1. Hitung SSE untuk kluster $\kappa! = j$ jika kasus disertakan.
 2. Jika SSE mengelompokkan kluster $\kappa <$ SSE kluster j , kasus akan berubah kluster.

Pada proses K-Means, nilai *within sum of squares* menunjukkan variasi kluster, semakin kecil nilai *within sum of squares* maka kluster yang terbentuk semakin baik (Calinski & Harabasz, 1974).

Optimalisasi jumlah kluster yang dibangun dapat diketahui dengan cara memeriksa apakah jumlah kluster tersebut telah optimum atau tidak. Ada beberapa macam teknik pemilihan jumlah kluster optimum. Dua diantaranya adalah Metode Elbow dan Metode PseudoF.

Metode Elbow

Metode Elbow merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah kluster terbaik dengan cara melihat persentase hasil perbandingan antara jumlah kluster yang akan membentuk siku pada suatu titik (Madhulatha, 2012). Metode ini memberikan ide dengan cara memilih nilai kluster dan kemudian menambah nilai kluster tersebut untuk dijadikan model data dalam penentuan kluster terbaik dengan melihat nilai SSE. Selain itu, persentase perhitungan yang dihasilkan menjadi pembandingan antara jumlah *kluster* yang ditambah. Hasil persentase yang berbeda dari setiap nilai *kluster* dapat ditunjukkan dengan menggunakan grafik sebagai sumber informasinya. Pada titik tertentu akan terjadi grafik penurunan secara drastis dengan sebuah lekukan yang disebut dengan kriteria siku (tangan/red). Nilai itu kemudian menjadi nilai *k* atau jumlah kluster yang terbaik (Bholowalia & Kumar, 2014).

Algoritma Metode Elbow dalam menentukan jumlah kluster pada K-Means (Bholowalia & Kumar, 2014), yakni:

1. Inisialisasi $\kappa = 1$.
2. Mulai.
3. Naikkan nilai κ .
4. Hitung hasil *sum of square error* dari tiap nilai κ .
5. Lihat hasil *sum of square error* dari nilai κ yang turun secara drastis.
6. Tetapkan nilai κ yang berbentuk siku tangan pada grafik yang didapatkan.
7. Selesai.

Metode Pseudo F

Tujuan untuk mengetahui jumlah kelompok optimum dapat dilakukan dengan menggunakan kriteria nilai *Pseudo Fstatistics*. Persamaan untuk menghitung nilai *Pseudo Fstatistics* diberikan pada persamaan (2.50).

$$Pseudo\ Fstatistics = \frac{(R^2 / (\kappa - 1))}{(1 - R^2 / (n - \kappa))} \quad (2.50)$$

dengan $R^2 = \frac{(SST - SSW)}{SST}$, $SST = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (x_{ijk} - \bar{x}_j)^2$, dan

$$SSW = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (x_{ijk} - \bar{x}_{jk})^2.$$

Keterangan:

SST (Sum Square Total) merupakan total jumlah dari kuadrat jarak sampel terhadap rata-rata keseluruhan. *SSW (Sum Square Within)* ialah total jumlah kuadrat jarak sampel terhadap rata-rata kelompoknya, n merupakan banyaknya sampel, c adalah banyaknya variabel, dan p merupakan banyaknya kelompok. Variabel x_{ijk} merupakan sampel ke- i pada variabel ke- j kelompok ke- κ , \bar{x}_j yaitu rata-rata seluruh sampel pada variabel ke- j , serta \bar{x}_{jk} adalah rata-rata sampel pada ke- j dan kelompok ke- κ . Nilai *Pseudo Fstatistics* tertinggi menunjukkan bahwa jumlah kelompok digunakan untuk memartisi data lebih optimal (Orpin & Kostylev, 2006).

2.10 Infrastruktur Surabaya

Peningkatan penduduk dari daerah ke Kota Surabaya dalam proses urbanisasi menelurkan masalah mulai dari kepadatan penduduk, sampah, pendidikan, kesehatan, transportasi, hingga ekonomi. Proses urbanisasi menggerakkan pembangunan kota dengan cepat baik secara sosial dan ekonomi. Hal ini menyulitkan penyediaan sarana dan prasarana untuk mengimbangi kehidupan perkotaan itu sendiri. Salah satunya adalah masalah infrastruktur dan lalulintas. Jenis infrastruktur yang dimaksud menurut Peraturan Presiden Republik Indonesia Nomor 38 Tahun 2015 tentang Kerjasama Pemerintah dengan Badan Usaha dalam

Penyediaan Infrastruktur, ialah infrastruktur ekonomi dan infrastruktur sosial mencakup:

- a. Infrastruktur transportasi;
- b. Infrastruktur jalan;
- c. Infrastruktur sumber daya air dan irigasi;
- d. Infrastruktur air minum;
- e. Infrastruktur sistem pengelolaan air limbah terpusat;
- f. Infrastruktur sistem pengelolaan persampahan;
- g. Infrastruktur telekomunikasi dan informatika;
- h. Infrastruktur ketenagalistrikan;
- i. Infrastruktur minyak dan gas bumi dan energi terbarukan;
- j. Infrastruktur konservasi energi;
- k. Infrastruktur fasilitas perkotaan;
- l. Infrastruktur fasilitas pendidikan;
- m. Infrastruktur fasilitas sarana dan prasarana olahraga, serta kesenian;
- n. Infrastruktur kawasan;
- o. Infrastruktur pariwisata;
- p. Infrastruktur lembaga permasyarakatan; dan
- q. Infrastruktur perumahan rakyat.

Penelaah yang dilakukan terhadap permasalahan pembangunan level lokal dan regional menimbulkan isu strategis yang terlansir dalam Laporan Kinerja Pemerintah Kota Surabaya Tahun 2016 berupa peningkatan infrastruktur untuk mendukung peran dan daya saing Kota Surabaya dalam menghadapi persaingan global yang melingkupi:

- a. Pembangunan dan pengembangan sistem transportasi, khususnya pada aspek pengembangan sistem angkutan massal
- b. Penyediaan dan penuntasan jaringan jalan
- c. Peningkatan kapasitas jalan secara memadai

Adapun aspek-aspek yang perlu dipertimbangkan di dalam perencanaan pengendalian fasilitas infrastruktur Kota Surabaya seperti yang terlansir pada tulisan oleh Tri Rismaharini (2013) diantaranya ialah pengendalian terhadap penggunaan lahan dan pengendalian garis sempadan jalan/bangunan khususnya di

sepanjang jalanan utama kota serta jalanan-jalanan lainnya yang dianggap strategis terhadap sistem transportasi kawasan.

(Halaman ini sengaja dikosongkan)

BAB 3

METODOLOGI PENELITIAN

3.1 Data dan Sumber Data

Penelitian yang dilakukan ialah analisis sentimen terhadap data Twitter mengenai pembangunan infrastruktur Kota Surabaya. Data bersumber dari warga dan atau berkaitan dengan Kota Surabaya, maka data diambil dengan kata pencarian ke akun-akun Twitter yang bergerak dibawah Kota Surabaya. Akun-akun tersebut ialah @sapawargaSby dan @e100ss. Akun @sapawargaSby merupakan salah satu akun media sosial resmi milik pemerintahan kota, sementara @e100ss adalah akun Twitter milik Radio Surabaya di frekuensi FM 100.00. Radio Surabaya ialah radio FM paling populer di masyarakat Surabaya. Radio ini mayoritas selalu dibunyikan warga Surabaya untuk memantau kemacetan yang terjadi di ruas jalan Surabaya dan juga melaporkan kejadian-kejadian yang bisa diantisipasi.

Pengambilan data dilakukan dengan proses khusus menggunakan Twitter *Application Programming Interface* (API). Data yang diambil berbahasa Indonesia. Pada Twitter API, pengguna dapat mengakses data *tweet* untuk mengolah dan menganalisis informasi dari Twitter tersebut. Proses yang dimaksud ialah membuat *Consumer Key*, *Consumer Secret*, *Access Token*, dan *Access Token Secret* terlebih dahulu yang akan berfungsi sebagai kunci agar aplikasi yang akan dibuat dapat diketahui oleh Twitter. Setelah mendapatkan keempat kunci di atas, selanjutnya ambil data *tweet* lalu disimpan ke dalam bentuk Excel format *Comma Separated Value* (CSV).

Data *tweet* dikumpulkan sejak 29 Agustus hingga Oktober 2017 dan Januari hingga 7 Februari 2018. Data yang terkumpul totalnya sebanyak 84.857 *tweet*. Namun, data mentah tersebut mengalami proses pemilahan secara manual berdasarkan kandungan sentimen terhadap infrastruktur Kota Surabaya sehingga didapatkan data olah sebanyak 1.500 data.

3.2 Struktur Data

Data dari setiap pencarian nama akun digabung menjadi 1 file. Struktur data yang siap diolah ditunjukkan oleh Tabel 3.1

Tabel 3.1 Struktur Data

<i>Tweet ke-</i>	$x_1 =$ Term 1	$x_2 =$ Term 2	...	$x_p =$ Term P
1	frekuensi x_1 pada <i>tweet ke-1</i>	frekuensi x_2 pada <i>tweet ke-1</i>	...	frekuensi x_p pada <i>tweet ke-1</i>
2	frekuensi x_1 pada <i>tweet ke-2</i>	frekuensi x_2 pada <i>tweet ke-2</i>	...	frekuensi x_p pada <i>tweet ke-2</i>
⋮	⋮	⋮	⋮	⋮
n	frekuensi x_1 pada <i>tweet ke-n</i>	frekuensi x_2 pada <i>tweet ke-n</i>	...	frekuensi x_p pada <i>tweet ke-n</i>

Data mentah diklasifikasikan secara manual dan subjektif menurut konteks yang membahas infrastruktur Kota Surabaya dan bentuk sentimennya: keluhan atau apresiasi. Contoh pengklasifikasian ditunjukkan pada Tabel 3.2.

Tabel 3.2 Contoh Bentuk Data

No.	Teks	Klasifikasi
1.	Semoga dengan perkembangan yang selama ini kita raih dapat berguna bagi kita bersama untuk kota tercinta kita... https://t.co/loJWhRs3lP	Positif
⋮	⋮	⋮
i.	Perlu diadakan lg swiping pelajar di kawasan surabaya barat,krn banyak yg membolos dan berada di warung kopi Tiap pagi @e100ss,@satpolppsby	Negatif
⋮	⋮	⋮
n.	@satpolppsby Iya min trims. Ditunggu untuk normalisasi trotoar n bahu jalannya agar pengguna jalan kaki dan motor lancar. @OmbudsmanRI137	Positif

3.3 Langkah Analisis

Langkah-langkah analisis yang akan dilakukan dalam penelitian ini, dipaparkan sebagai berikut.

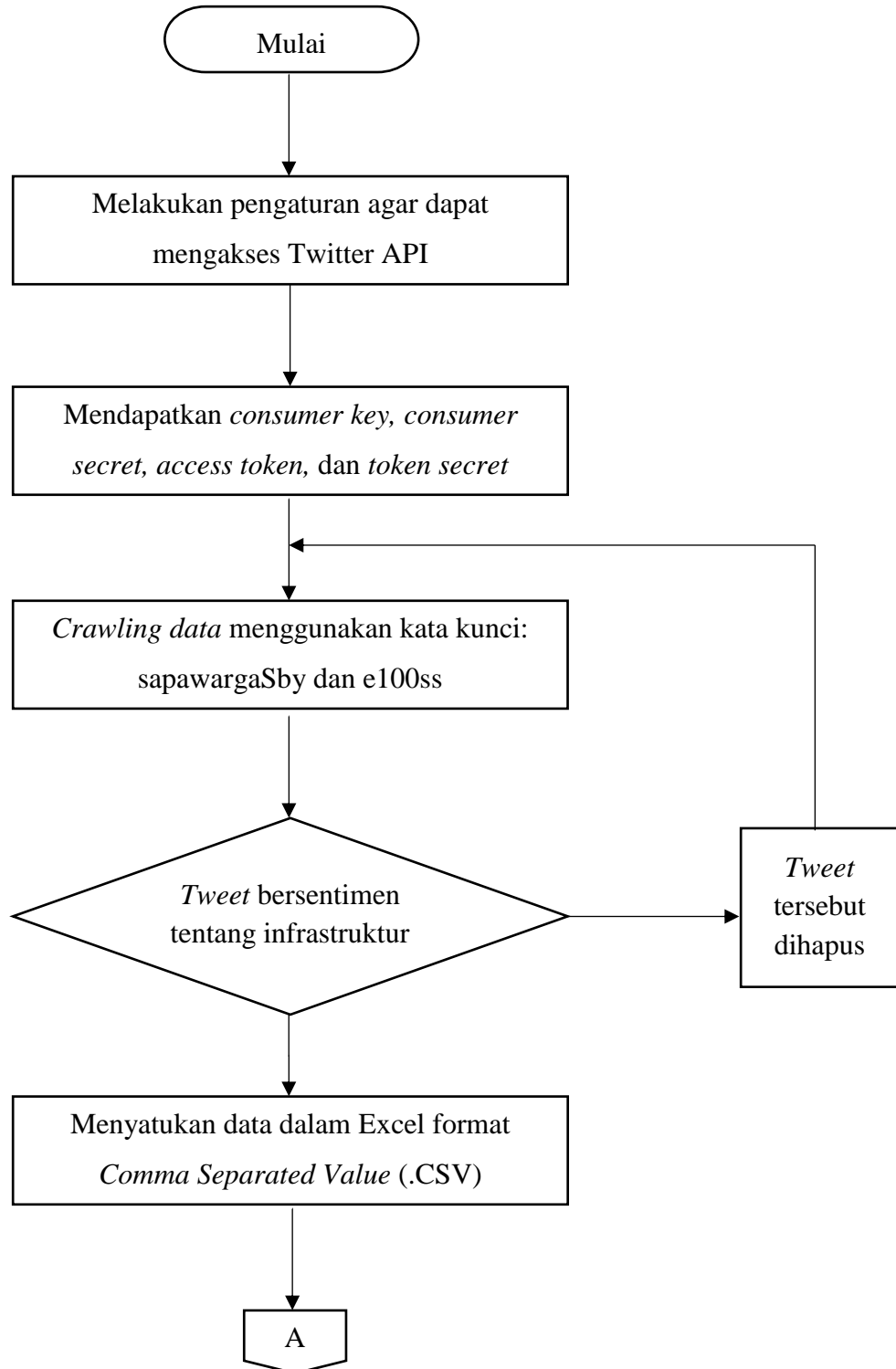
1. Ambil data *tweet* melalui Twitter API.
 - a. Buka laman <https://appstwitter.com> lalu *sign in* akun Twitter pribadi.
 - b. Buat sebuah aplikasi API dengan nama dan tujuan yang jelas. Hal ini untuk memberi penjelasan pada Twitter perihal untuk apa pengambilan data tersebut dilakukan.
 - c. Lakukan pengaturan dan dapatkan kode *Consumer Key*, *Consumer Secret*, *Access Token*, dan *Access Token Secret* yang digunakan sebagai syarat kunci pengambilan data.
 - d. Masukkan *keyword* yang mengandung *sapawargaSby*, dan *e100ss*.
 - e. Simpan hasil *crawling* ke dalam dataset dengan format Excel.CSV.
2. Melakukan *preprocessing* sebagaimana *preprocessing* diperlukan untuk menghindari data yang belum siap olah, seperti: data yang kurang sempurna, terdapat gangguan (*noise*) pada data, dan data yang tidak konsisten (Hemalatha, 2012). Tahapan praproses pada data teks ini yaitu:
 - a. Menghapus data *tweet* yang tidak mengandung sentimen, dan juga yang mengandung dua sentimen sekaligus.
 - b. Menghapus tulisan “RT” yang merupakan simbol dari *retweet* atau *response tweet*.
 - c. Menghapus link URL.
 - d. Menghapus simbol dan tanda baca, seperti: `~!@#%&*()_+ = { } [] \ ; : , < > , .
 - e. Lakukan *case folding* berupa menyeragamkan setiap teks menjadi huruf non kapital.
 - f. Menghapus kata pada teks *tweet* yang terkandung dalam *stopwords*.
 - g. Lakukan *tokenizing* yaitu pemecahan kalimat *tweet* menjadi kata per kata.
3. Membuat *document term matrix* dan memberikan pembobotan TF-IDF.
4. Klasifikasi data sentimen menggunakan metode *Support Vector Machine* (SVM) dan metode *Neural Network*.

- a. Menentukan pembobot parameter pada SVM dan membangun model SVM.
 - b. Melakukan pengerjaan pelatihan *Neural Network* menurut algoritma *Backpropagation*.
 - c. Evaluasi hasil klasifikasi. Evaluasi dilakukan dengan menghitung ketepatan klasifikasi berdasarkan tingkat akurasi, presisi, *recall*, dan skor F1.
5. Membentuk klaster berdasarkan kata atau *term* yang muncul pada *tweet* sentimen negatif dan positif untuk mengetahui topik *term* yang spesifik dibahas dari *tweet* masyarakat Surabaya.
- a. Memisahkan data bersentimen negatif dan data yang bersentimen positif dari data awal.
 - b. Menerapkan algoritma pembentukan klaster K-means untuk mengelompokkan kata atau *term*.
 - c. Melakukan evaluasi terhadap banyaknya klaster sebagai optimalisasi jumlah klaster.
 - d. Memeriksa kata terbanyak di setiap klaster yang membentuk sebuah frasa *term* spesifik yang sering dibahas oleh masyarakat Surabaya di Twitter.
6. Interpretasi dan menarik kesimpulan.

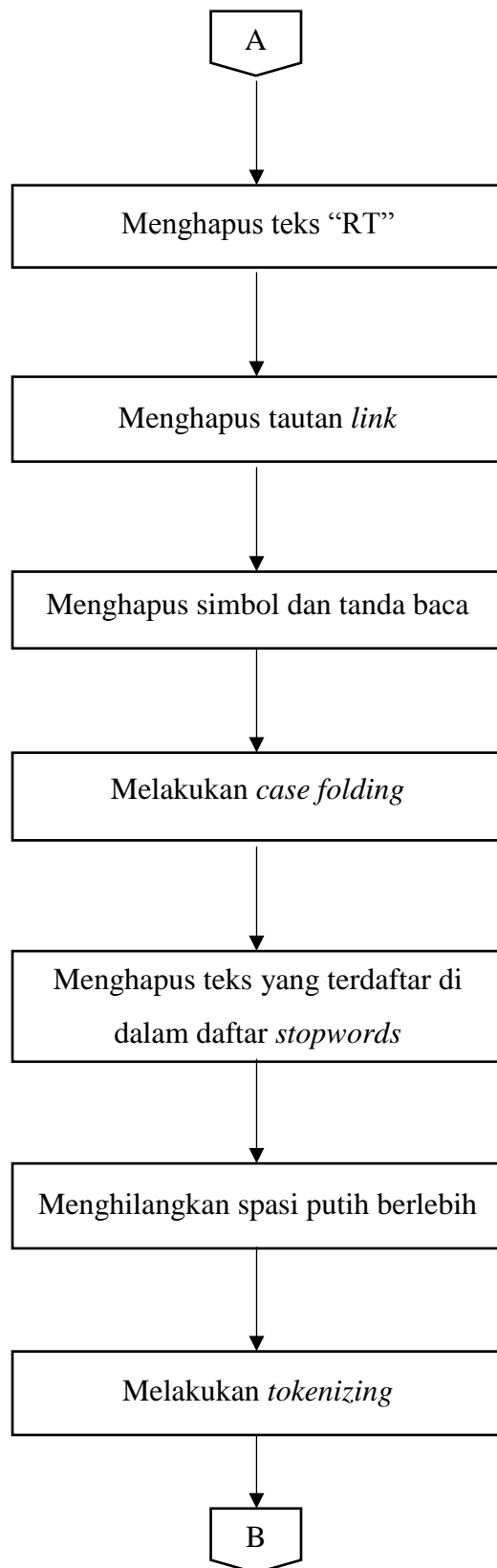
3.4 Diagram Alir

Skema diagram alir penelitian yang dilakukan ini disajikan pada Gambar

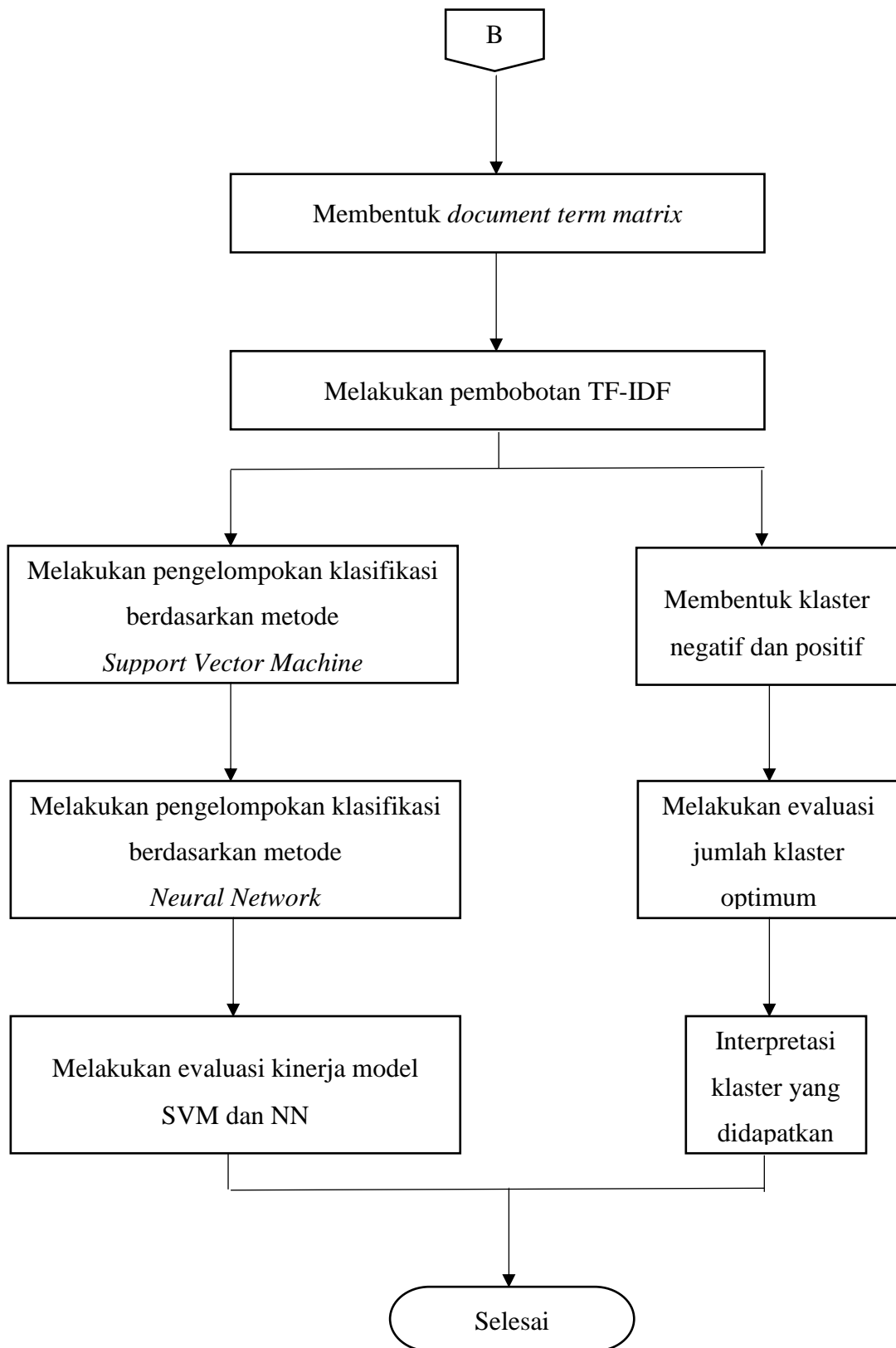
3.1.



Gambar 3.1 Diagram Alir Penelitian



Gambar 3.1 Diagram Alir Penelitian (lanjutan)



Gambar 3.1 Diagram Alir Penelitian (lanjutan)

(Halaman ini sengaja dikosongkan)

BAB 4

ANALISIS DAN PEMBAHASAN

Sentimen dalam penelitian ini pada dasarnya mengacu pada polaritas kontekstual dari teks atau dokumen yang mengandung efek emosional pada pembaca yang ingin disampaikan oleh penulis tentang subjek atau objek yang berkaitan dengan pembangunan infrastruktur Kota Surabaya. Sentimen bersifat subjektif dan bergantung pada moral pribadi, nilai-nilai moral, dan keyakinan seseorang. Dalam konteks tertentu, seseorang pada dasarnya cerdas dalam menilai sentimen. Namun, memerintah komputer untuk melakukan hal yang sama tidaklah mudah. Ilmu komputer memiliki bidang khusus untuk hal tersebut yang melibatkan bahasa manusia yang disebut Pemrosesan Bahasa Alami (*Neuro-Linguistic Programming*) atau NLP. Area yang berhubungan dengan penggalian informasi subjektif dari beberapa teks atau dokumen disebut Analisis Sentimen.

Kalimat atau kata-kata positif memiliki sentimen positif yang melekat pada teks atau dokumen. Sebagai contoh, ketika beberapa teks menunjukkan kebahagiaan, antusiasme, kebaikan, dll. Demikian pula kalimat atau kata-kata negatif mempunyai sentimen negatif yang melekat pada teks atau dokumen. Sebagai contoh, ketika teks tersebut menunjukkan kesedihan, kebencian, kekerasan, diskriminasi, dll. Apabila tidak ada emosi yang tersirat, maka teks atau dokumen tersebut tergolong netral.

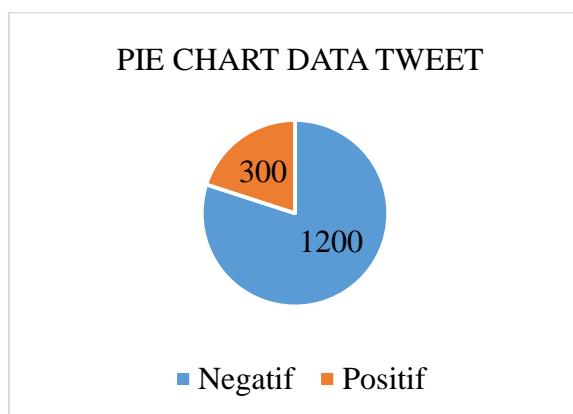
4.1 Klasifikasi Sentimen Data Twitter Menggunakan Metode SVM dan Neural Network.

Pada subbab ini akan dipaparkan secara ringkas mengenai karakteristik data *tweet* yang akan diolah, langkah-langkah praproses yang dilakukan agar data siap diolah, dan pengolahan klasifikasi menggunakan dua metode berbeda yaitu *Support Vector Machine* dan *Neural Network*.

4.1.1 Karakteristik Data

Tahap awal yang dilakukan di setiap penelitian adalah statistika deskriptif, yaitu melihat eksplorasi data secara menyeluruh. Hal ini dilakukan untuk melihat

gambaran umum dari data tersebut. Data diekstraksi dari akun resmi Twitter pemerintahan kota @sapawargaSby dan akun resmi Twitter Radio Surabaya 100.00 FM bernama @e100ss. Data *tweet* yang mengandung sentimen tentang infrastruktur terkumpul sebanyak 1500 data. Berdasarkan jenis sentimen tersebut, data terbagi menjadi dua tipe yaitu sentimen negatif dan sentimen positif. Sentimen negatif berjumlah 1200 data dan sentimen positif berjumlah 300 data. Gambar 4.1 menampilkan sebuah *pie chart* yang memvisualisasikan proporsi data.



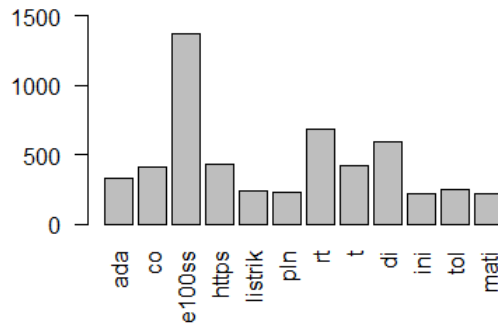
Gambar 4.1 *Pie Chart Data Tweet* Berdasarkan Kandungan Sentimen

Berdasarkan informasi yang didapatkan, diketahui bahwa *tweet* yang memiliki sentimen negatif jauh lebih banyak dibandingkan dengan *tweet* yang sentimen positif. Hal ini dikarenakan data yang dikumpulkan dari Twitter lebih banyak mayoritas mengenai keluhan dan pengaduan mengenai hal-hal yang dirasa mengganggu, menghambat aktivitas, atau membahayakan jiwa. Contoh keluhan dan pengaduan yang dimaksud ditampilkan pada Tabel 4.1.

Tabel 4.1 Contoh Data Awal

No	<i>Tweet</i>	Class	Jumlah Karakter
1	Delay di bandara juanda..shrsnya brgkt jm18.35 smpe skrng blm ada tanda2 mw diberangkatkan..@e100ss @kemenhub151 https://t.co/iELU1iDepK	Negatif	135
2	Late post @e100ss enaknya ada jalur baru khusus motor di tambaksari, masih baru selesai lho pemasangan keramiknya <ed><U+00A0><U+00BE><ed><U+00B4><U+00A3> https://t.co/o3WvyyMZke	Positif	178

Kata-kata *term* yang terkumpul pada 1500 *tweet* secara total berjumlah 3055 kata. Secara garis besar, frekuensi data awal yang dikumpulkan ditampilkan pada Gambar 4.2.



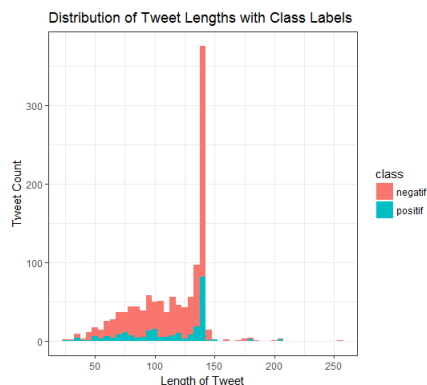
Gambar 4.2 Histogram Frekuensi Data Awal

Berdasarkan Gambar 4.2, frekuensi kata tertinggi di miliki oleh kata “e100ss”. Dapat dikatakan kata “e100ss” sendiri berjumlah 1373 kata di dalam 1500 *tweet* dataset. Selain itu bisa juga diketahui ringkasan statistika deskriptif persebaran jumlah karakter pada data *tweet* seperti yang ditunjukkan pada Tabel 4.2.

Tabel 4.2. Statistika Deskriptif

Jumlah <i>Tweet</i>	Min.	Median	Mean	Maks.	Modus	Standar Deviasi
1500	23,0	127,0	115.6	253.0	140.0	29,595

Visualisasi banyak karakter per-*tweet* berdasarkan label kelas positif negatif ditampilkan pada Gambar 4.3.



Gambar 4.3 Distribusi Panjang *Tweet* dan Label Kelasnya

4.1.2 Perangkat Pengujian

Kriteria perangkat pengujian yang digunakan dalam penelitian ini terdiri dari perangkat keras dan perangkat lunak, seperti yang ditunjukkan berturut-turut pada Tabel 4.3.

Tabel 4.3. Spesifikasi *Hardware* dan *Software*

Perangkat	Spesifikasi
Jenis	Laptop
Processor	Intel(R) Core (TM) i5-4210U CPU @ 1.70GHz 2.40 GHz
Memori (RAM)	8 GB
Sistem Operasi	Windows 8
Tools	Rstudio R 3.4 Python 3.5.5 Weka 3.8

4.1.3 Preprocessing Data

Tujuan dilakukan *preprocessing* terhadap data ialah untuk menghilangkan *noise*, memperjelas fitur, dan meng-konversi data asli agar data yang diolah sesuai dengan kebutuhan penelitian. Tahapan *preprocessing* terdiri dari:

1. Menghapus data *tweet* yang topiknya bukan mengenai infrastruktur, data *tweet* yang tidak mengandung sentimen, dan juga data *tweet* yang mengandung dua sentimen sekaligus.
2. Menghapus tautan *link* yang terdapat di dalam *tweet*. Istilah "*link*" atau "*hyperlink*" adalah tautan yang menghubungkan satu halaman ke halaman lain di suatu *blog* atau *website* (*link* internal) maupun menghubungkan halaman suatu *blog* atau *website* ke halaman situs *blog* atau *website* yang lain (*link* eksternal). Biasanya *link* ditemukan di sejumlah *tweet* yang mengarahkan pengguna atau *user* lain yang membaca *tweet* tersebut untuk melihat informasi tambahan yang ditaruh di halaman yang lain.

No	Input	Output
1	RT @setyobudianto76: @e100ss Jagir Sidomukti listrik padam padahal ada acara Bazar kampung nih..PLN gimana? https://t.co/CNwx8XAPCg	RT @setyobudianto76: @e100ss Jagir Sidomukti listrik padam padahal ada acara Bazar kampung nih..PLN gimana?
2	@Effendi07Joni @pln_123 @e100ss @RadioWarna Di daerah sidokepung sekitarnya koq juga padam ada apa ya..koq slalu ha... https://t.co/sw3DKAGqGs	@Effendi07Joni @pln_123 @e100ss @RadioWarna Di daerah sidokepung sekitarnya koq juga padam ada apa ya..koq slalu ha...
3	RT @Syam123A: @e100ss Assalamualaikum, ada masalah telp dan internet kami sampai 3 hari ini tdk bisa conek, dg no telp (031) 99035677	RT @Syam123A: @e100ss Assalamualaikum, ada masalah telp dan internet kami sampai 3 hari ini tdk bisa conek, dg no telp (031) 99035677

3. Menghapus tulisan “RT” yang merupakan singkatan dari *retweet* atau *reposting tweet*. Kegunaan dari “RT” sendiri ialah untuk membantu pengguna Twitter untuk mempublis ulang dan membagi suatu *tweet* secara cepat kepada *followers* pengguna.

No	Input	Output
1	RT @setyobudianto76: @e100ss Jagir Sidomukti listrik padam padahal ada acara Bazar kampung nih..PLN gimana?	@setyobudianto76: @e100ss Jagir Sidomukti listrik padam padahal ada acara Bazar kampung nih..PLN gimana?
2	@Effendi07Joni @pln_123 @e100ss @RadioWarna Di daerah sidokepung sekitarnya koq juga padam ada apa ya..koq slalu ha...	@Effendi07Joni @pln_123 @e100ss @RadioWarna Di daerah sidokepung sekitarnya koq juga padam ada apa ya..koq slalu ha...
3	RT @Syam123A: @e100ss Assalamualaikum, ada masalah telp dan internet kami sampai 3 hari ini tdk bisa conek, dg no telp (031) 99035677	@Syam123A: @e100ss Assalamualaikum, ada masalah telp dan internet kami sampai 3 hari ini tdk bisa conek, dg no telp (031) 99035677

4. Menghapus *username*. Maksud dari *username* pada Twitter adalah identitas pengguna.

No	Input	Output
1	@setyobudianto76: @e100ss Jagir Sidomukti listrik padam padahal ada acara Bazar kampung nih..PLN gimana?	: Jagir Sidomukti listrik padam padahal ada acara Bazar kampung nih..PLN gimana?
2	@Effendi07Joni @pln_123 @e100ss @RadioWarna Di daerah sidokepung sekitarnya koq juga padam ada apa ya..koq slalu ha...	Di daerah sidokepung sekitarnya koq juga padam ada apa ya..koq slalu ha...
3	@Syam123A: @e100ss Assalamualaikum, ada masalah telp dan internet kami sampai 3 hari ini tdk bisa conek, dg no telp (031) 99035677	Assalamualaikum, ada masalah telp dan internet kami sampai 3 hari ini tdk bisa conek, dg no telp (031) 99035677

5. Menghapus tanda baca, seperti `~!@#\$\$%^&*()_+={ }[]\|:;''><,. Tanda baca adalah simbol yang tidak berhubungan dengan fonem (suara) atau kata dan frasa pada suatu bahasa, melainkan berperan untuk menunjukkan struktur dan organisasi tulisan, dan juga intonasi serta jeda yang dapat diamati sewaktu pembacaan. Selain pada tulisan kata, tanda baca juga ditemukan di tautan *link*. Beberapa aspek tanda baca adalah suatu gaya spesifik yang karenanya tergantung pada pilihan penulis.

No	Input	Output
1	: Jagir Sidomukti listrik padam padahal ada acara Bazar kampung nih..PLN gimana?	Jagir Sidomukti listrik padam padahal ada acara Bazar kampung nihPLN gimana
2	Di daerah sidokepung sekitarnya koq juga padam ada apa ya..koq slalu ha...	Di daerah sidokepung sekitarnya koq juga padam ada apa yakoq slalu ha...
3	Assalamualaikum, ada masalah telp dan internet kami sampai 3 hari ini tdk bisa conek, dg no telp (031) 99035677	Assalamualaikum ada masalah telp dan internet kami sampai 3 hari ini tdk bisa conek dg no telp 031 99035677

6. Melakukan *case folding* berupa menyeragamkan setiap teks menjadi huruf non kapital.

No	Input	Output
1	Jagir Sidomukti listrik padam padahal ada acara Bazar kampung nihPLN gimana	jagir sidomukti listrik padam padahal ada acara bazar kampung nihpln gimana
2	Di daerah sidokepung sekitarnya koq juga padam ada apa yakoq slalu ha...	di daerah sidokepung sekitarnya koq juga padam ada apa yakoq slalu ha...
3	Assalamualaikum ada masalah telp dan internet kami sampai 3 hari ini tdk bisa conek dg no telp 031 99035677	assalamualaikum ada masalah telp dan internet kami sampai 3 hari ini tdk bisa conek dg no telp 031 99035677

7. Menghapus kata pada *tweet* yang terkandung di dalam *stopwords*. Daftar kata *stopwords* adalah daftar kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Fungsi dari penghapusan ini adalah untuk mengurangi jumlah kata yang akan diproses.

No	Input	Output
1	jagir sidomukti listrik padam padahal ada acara bazar kampung nihpln gimana	jagir sidomukti listrik padam acara bazar kampung nihpln gimana
2	di daerah sidokepung sekitarnya koq juga padam ada apa yakoq slalu ha...	daerah sidokepung koq padam yakoq slalu ha...
3	assalamualaikum ada masalah telp dan internet kami sampai 3 hari ini tdk bisa conek dg no telp 031 99035677	assalamualaikum telp internet 3 conek dg no telp 031 99035677

8. Menghapus digit angka. Digit atau angka adalah sebuah simbol untuk penomoran yang digunakan pada bilangan. Kata digit berasal dari bahasa latin kuno, yaitu *digita*, yang berarti 10 jari tangan manusia yang berhubungan dengan sistem bilangan basis 10. Sepuluh digit angka diurutkan berdasarkan nilainya adalah 0,1,2,3,4,5,6,7,8,dan 9.

No	Input	Output
1	jagir sidomukti listrik padam acara bazar kampung nihpln gimana	jagir sidomukti listrik padam acara bazar kampung nihpln gimana
2	daerah sidokepong koq padam yakoq slalu ha...	daerah sidokepong koq padam yakoq slalu ha...
3	assalamualaikum telp internet 3 conek dg no telp 031 99035677	assalamualaikum telp internet conek dg no telp

9. Menghilangkan *strip white space*. Karakter *white space* yang dimaksud yakni: spasi, tab, baris baru, dan lain-lain yang sejenis.

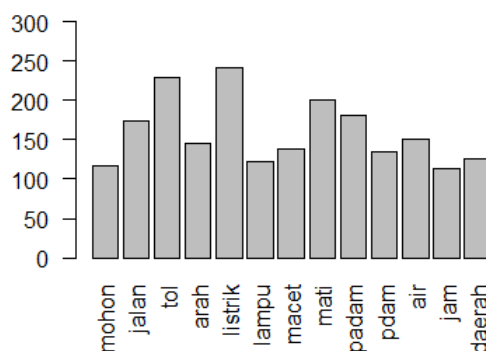
No	Input	Output
1	jagir sidomukti listrik padam acara bazar kampung nihpln gimana	jagir sidomukti listrik padam acara bazar kampung nihpln gimana
2	daerah sidokepong koq padam yakoq slalu ha...	daerah sidokepong koq padam yakoq slalu ha...
3	assalamualaikum telp internet conek dg no telp	assalamualaikum telp internet conek dg no telp

10. Melakukan *tokenizing*.

No	Output
1	“jagir”, “sidomukti”, “listrik”, “padam”, “acara”, “bazar”, “kampung”, “nihpln”, “gimana”
2	“daerah”, “sidokepong”, “koq”, “padam”, “yakoq”, “slalu”, “ha...”
3	“assalamualaikum”, “telp”, “internet”, “conek”, “dg”, “no” “telp”

Berdasarkan praproses yang telah dilakukan, didapatkan total jumlah istilah sebanyak 2.824 istilah pada 1500 *tweet*. Istilah-istilah tersebut ditampilkan pada Lampiran 1.

Kata-kata *term* yang terkumpul pada 1500 *tweet* yang sudah mengalami praproses berjumlah 2824 kata. Secara garis besar, frekuensi data setelah mengalami praproses ditampilkan pada Gambar 4.4.



Gambar 4.4. Histogram Frekuensi Data Setelah Praproses

Berdasarkan Gambar 4.4, frekuensi kata tertinggi di miliki oleh kata “listrik”. Dapat dikatakan kata “listrik” berjumlah 241 kata di dalam 1500 *tweet* dataset. Posisi frekuensi tertinggi berikutnya diikuti oleh kata “tol” sebanyak 229 kata dan kata “padam” sebanyak 181 kata.

4.1.4 Document Term Matrix

Langkah setelah melakukan *preprocessing* adalah membuat *document term matrix* (DTM). Hal ini penting dilakukan pada koleksi data (*corpus*) yang tengah dikerjakan. DTM adalah sebuah matriks dua dimensi berukuran $n \times p$ dengan dokumen sebagai baris dan kata (*term*) sebagai kolom yang elemennya adalah jumlah kata di dalam satu dokumen. Fungsi DTM ialah untuk dapat mewakili topik dari satu dokumen dengan adanya frekuensi kata yang signifikan.

Awalnya di dalam matriks yang terbentuk, terdapat 4.223.433 sel yang berisi nol, ada 12.567 sel yang berisi tidak nol, dan 99,7% mayoritas isi sel-sel di dalam matriks tersebut ialah nol. Hal ini diketahui dari banyak sel berisi nol dibagi dengan total banyak sel berisi nol dan yang tidak nol. Kelangkaan ini kemudian dikurangi menggunakan teknik komputasi yang bernama *RemoveSparseTerm*. Tujuannya adalah menghapus istilah yang jarang ditemukan di dalam matriks DTM.

Jumlah term tersebut kemudian dikurangi dengan menggunakan *RemoveSparseTerm* dengan tujuan mereduksi kata yang jarang ditemukan di dalam *corpus*. Kejaringan ini diatur sebesar 99%, agar kata yang 99% berisi nol di dalam matriks akan dihapus. Hasilnya, tersisa sejumlah 110 kata yang berhubungan dengan infrastruktur Kota Surabaya. Matriks DTM 110 kata pada data *tweet* pendapat masyarakat terhadap pembangunan infrastruktur Kota Surabaya ditunjukkan pada Tabel 4.4.

Tabel 4.4. Document Term Matrix 110 Kata

Terms Docs	Listrik	Padam	kota	parkir	daerah	Internet	...	Hati
1	1	1	0	0	0	0		0
⋮	⋮	⋮	⋮	⋮	⋮	⋮		⋮
12	1	1	0	0	0	0		0
13	0	0	0	0	0	0		0
14	0	0	0	1	0	0		0
15	1	0	0	0	0	0		0
16	0	0	0	0	0	0		0
17	1	1	0	0	0	0		0
18	0	0	0	0	1	0		0
⋮								⋮
1500	0	0	0	0	0	0	...	1

4.1.5 Term-Weighting TF-IDF

Evaluasi untuk mengetahui seberapa penting suatu kata atau *term* pada koleksi dokumen atau *corpus* juga merupakan salah satu tahap dalam *text mining* dan *information retrieval*. Hal tersebut dititikberatkan pada semakin sering kata dalam dokumen muncul di dalam *corpus*. Teknik reduksi variabel yang dilakukan pada penelitian ini adalah *term frequency – inverse document frequency*, yang sering disebut dengan TF-IDF. Nilai TF-IDF meningkat secara proporsional sesuai dengan berapa kali suatu kata muncul pada dokumen dan diimbangi oleh frekuensi

kata dalam *corpus*. Hal ini bertujuan untuk menyesuaikan fakta bahwa ada beberapa kata yang muncul lebih sering di dalam *corpus* tersebut secara umum.

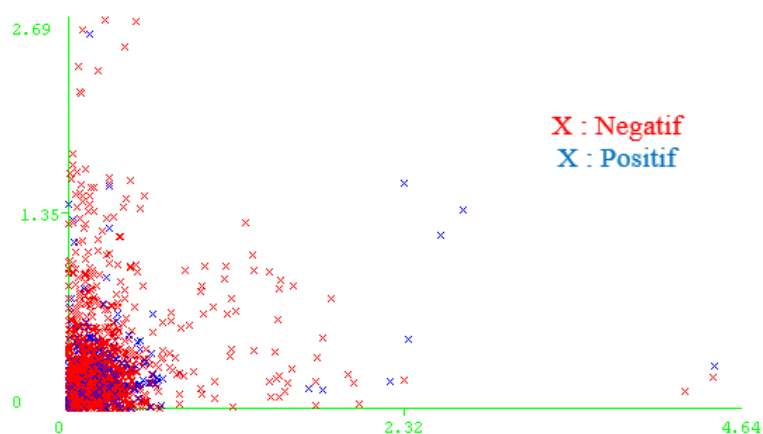
Tabel 4.5. Matriks DTM dengan pembobotan TF-IDF

Terms Docs	Listrik	Padam	...	Hati
1	1,346	1,533	...	0
⋮	⋮	⋮	⋮	⋮
12	1,346	1,533	...	0
13	0	0	...	0
14	0	0	...	0
15	1,346	0	...	0
16	0	0	...	0
17	1,346	1,533	...	0
18	0	0	...	0
⋮	⋮	⋮	⋮	⋮
1500	0	0	...	1,506

4.1.6 Klasifikasi dengan SVM

Data yang telah melalui tahap pra-proses kemudian dilatih dengan metode SVM menggunakan *10 fold cross validation*. Total data adalah 1.500 *tweet*, 9 *fold* berjumlah 1.350 data untuk *training* dan satu *fold* berjumlah 150 data untuk *testing*. Teknik sampling yang digunakan adalah *stratified cross validation* karena data merupakan data *imbalanced* dengan proporsi negatif berbanding positif ialah 80:20. Tahap ini bertujuan untuk mengklasifikasi data *training* terhadap dua kelas yaitu negatif dan positif. Proses ini membutuhkan model SVM dari hasil *training* data. Model SVM yang didapatkan kemudian diujikan pada data *testing*. Hasil dari model tersebut dapat diketahui dengan melakukan evaluasi lagi berdasarkan nilai akurasi, presisi, *recall*, dan skor F1.

Metode yang paling intuitif untuk mengidentifikasi apakah sampel dapat dipisahkan secara linear adalah dengan membuat plot sampel. Tentu saja, tidak mungkin dapat memvisualisasikan sampel jika dimensi ruang fitur lebih dari tiga. Sehingga, untuk sekedar memberikan contoh plot data, dibuatkan plot antara kata “listrik” dan “nyala” untuk melihat pola linearitas data *tweet* mengenai infrastruktur Kota Surabaya yang dipaparkan pada Gambar 4.5.



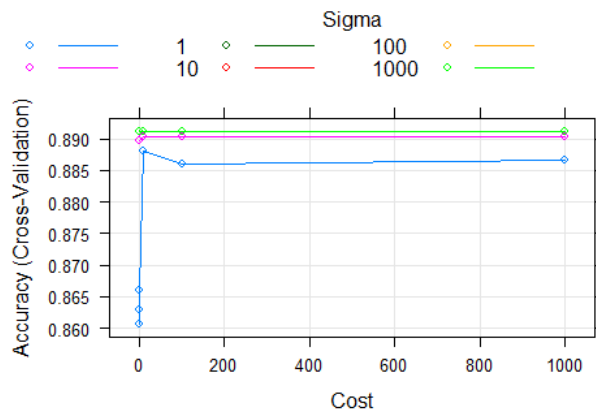
Gambar 4.5 Scatter Plot Variabel Listrik dan Nyala

Berdasarkan Gambar 4.5, terlihat bahwa data *tweet* tidak terpisah secara linier. Maka data *tweet* tersebut bersifat non-linier. Data non-linier dapat diklasifikasikan dengan SVM Non-linier.

Jika diperhatikan, pemisahan dua kelas ini berasal dari sampel-sampel yang lebih dekat dengan golongan kelas yang lain yaitu sampel-sampel yang berada pada *margin*. Perilaku ini menimbulkan kasus bahwa jumlah sampel yang digunakan untuk mendapatkan garis pemisah yang optimal dari sudut pandang data latihan yang digunakan (*overfitting*) untuk sebagian besar data. Pemisahan akan optimal dari sudut pandang data *training*, tetapi saat digeneralisasi akan menjadi buruk untuk data yang bukan data *training*. Oleh sebab itu, untuk menghindari *overfitting* dilakukan modifikasi pada masalah optimasi untuk mengoptimalkan kecocokan garis ke data dan memberikan hukuman pada sejumlah sampel di dalam *margin* pada saat yang bersamaan. Di sini C didefinisikan sebagai penalti dari berapa banyak sampel di dalam *margin* yang berkontribusi pada kesalahan keseluruhan.

Akibatnya, dengan nilai C , peneliti dapat menyesuaikan klasifikasi yang seharusnya. Nilai C yang besar akan memberikan bias yang rendah dan varians yang tinggi. Nilai C kecil akan memberikan bias lebih tinggi dan varians yang lebih rendah.

Kasus non-linier dapat diatasi dengan adanya parameter gamma (γ). Titik-titik data Gambar 4.5 tidak dapat dipisahkan secara linier dalam dua dimensi sehingga perlu diubah dimensinya menjadi dimensi yang lebih tinggi dimana data akan dapat dipisahkan. Cara menaikkan dimensi tersebut ialah dengan cara menggunakan kernel *Radial Basis Function* (RBF). Parameter gamma mengontrol bentuk “puncak” setiap dilakukan kenaikan. Gamma bernilai kecil akan memberikan bias yang rendah dan varians yang tinggi sementara gamma bernilai besar akan memberikan bias yang lebih tinggi dan varians yang lebih rendah.



Gambar 4.6 Plot Pencarian Parameter C dan Sigma dengan Teknik *Grid Search*

Adapun penentuan nilai kedua parameter tersebut menggunakan teknik *Grid Search*. Pencarian grid atau *grid search* adalah proses pemindaian data untuk mengkonfigurasi parameter optimal untuk model yang diberikan. Parameter optimal yang didapatkan berdasarkan teknik *grid search* ialah 0,01 dan 1000 berturut-turut untuk C dan γ . Hasil pemindaian *grid search* dilampirkan Lampiran 2 dan ditampilkan pada Gambar 4.6.

Hasil ketepatan klasifikasi terbaik didapatkan dengan menggunakan parameter $C = 0,01$ dan $\gamma = 1000$. Nilai gamma sendiri kemudian disubstitusikan

pada persamaan kernel RBF (2.32) untuk mendapatkan fungsi kernel. Sehingga, fungsi Kernel RBF yang digunakan ialah sebagai berikut.

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-1000 \|\mathbf{x}_i - \mathbf{x}_j\|^2)$$

Kemudian fungsi kernel tersebut digunakan untuk membentuk mendapatkan fungsi *hyperplane* dengan cara mensubstitusikan fungsi kernel pada persamaan (2.29). Dengan demikian, fungsi *hyperplane* dari data *training* yang didapatkan yakni seperti berikut.

$$\hat{f}(x) = \sum_{i=1}^{849} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_j) - 0,996$$

Dengan α_i ialah *Lagrange multiplier* dari *support vector* yang berukuran (849,1) atau bisa juga disebut dengan vektor koefisien. Variabel y_i merupakan label kelas penanda kandungan sentimen yang nilainya -1 untuk sentimen negatif dan +1 untuk sentimen positif. Variabel x merupakan nilai input *term* dan nilai $-0,996$ adalah nilai bias. Nilai-nilai yang terkandung untuk masing-masing variabel ditampilkan pada Lampiran 3.

Model SVM yang didapatkan kemudian digunakan pada data *testing* untuk melihat ketepatan klasifikasi model. Ketepatan klasifikasi dapat diketahui dengan dilakukannya beberapa pengukuran evaluasi kinerja model. kriteria pengukuran ketepatan klasifikasi yang digunakan oleh peneliti pada data testing dipaparkan pada Tabel 4.6. Adapun untuk tabel *confusion matrix* yang diperoleh dilampirkan pada Lampiran 4.

Tabel 4.6. Ketepatan Klasifikasi Metode SVM Non-linier Kernel RBF C=0,01, Gamma=1000

Akurasi	Presisi	Recall	Skor F1
92,67%	66,67%	92,24%	77,39%

Informasi yang dapat diketahui dari Tabel 4.6 ialah kriteria pengukuran ketepatan klasifikasi adalah akurasi, presisi, *recall* dan skor F1. Metode SVM Non-linier dengan Kernel RBF menggunakan parameter C=0,01 dan $\gamma = 1000$

mengeluarkan nilai akurasi sebesar 92,67%. Nilai presisi, yaitu tingkat ketepatan antara informasi yang diminta oleh pengguna dengan hasil yang diberikan oleh sistem telah tepat sebesar 66,67%. Nilai *recall* sebesar 92,24% merupakan ketepatan sistem untuk mengklasifikasi data dengan benar. Adapun skor F1 yang merupakan pengukuran akurasi dataset yang dihitung dengan cara mengambil rata-rata harmonik dari presisi dan *recall* ialah sebesar 77,39%. Nilai-nilai tersebut dapat mewakili keakuratan model dalam mengklasifikasi data *tweet* sentimen warga Surabaya terhadap pembangunan infrastruktur.

4.1.7 Klasifikasi dengan Neural Network

Tahap ini bertujuan untuk mengklasifikasi data *training* terhadap dua kelas yaitu negatif dan positif. Proses ini membutuhkan model Neural Network dari hasil data *training*. Model *Neural Network* dibangun menggunakan algoritma *backpropagation* dengan 110 jumlah input, 1 *hidden layer*, 71 jumlah neuron, dan 2 *output layer* yaitu kelas negatif dan kelas positif. Secara sederhana, penentuan struktur model *neural network* dapat dipahami melalui Tabel 4.7.

Cara yang ditempuh untuk menentukan jumlah neuron di dalam *hidden layer* adalah *trial-error* memasukkan jumlah neuron pada model yang dibentuk. Fungsi aktivasi yang digunakan adalah *sigmoid*. Penentuan dilakukan berdasarkan nilai akurasi yang bernilai besar yang mendekati 1. Nilai akurasi tersebut dikeluarkan dari *cross-validation* model yang dilakukan.

Akurasi yang paling baik dikeluarkan oleh model neural network dengan jumlah neuron sebanyak 71 neuron. Model tersebut melalui proses sampling dengan teknik stratifikasi 10 *fold cross validation* yang dilampirkan pada Lampiran 5. Model yang didapatkan memberikan nilai bobot yang dirumuskan oleh persamaan (2.34) dan (2.36). Bobot tersebut disubstitusikan pada persamaan fungsi sigmoid yang dirumuskan pada persamaan (2.35) dan (2.37) berurutan. Berdasarkan jaringan yang dibangun, model *Neural Network* satu *hidden layer* dengan 71 jumlah neuron dan menggunakan fungsi aktivasi sigmoid yang terbentuk di *hidden layer* pada setiap neuron dituliskan pada Tabel 4.8. Adapun bobot neuron yang berada di *hidden layer* dilampirkan pada Lampiran 6 sebagai ilustrasi *output* yang didapatkan.

Tabel 4.7. Ringkasan Struktur Model Neural Network

Layer	Jumlah Layer	Jumlah Neuron	Data
Input	1	110	listrik, padam, kota, parkir, daerah, internet, mohon, box, masuk, kabel, kebakaran, suramadu, tol, padat, pdam, pipa, proyek, bocor, jalan, kali, cctv, sultan, jalur, malam, sms, terima, berhenti, pratama, total, lampu, mati, timur, biar, masyarakat, pemkot, air, airnya, barat, macet, industri, satelit, arus, perbaikan, toll, dupak, lancar, ramai, babatan, indah, perumahan, jam, min, ini, jaya, etoll, gate, petugas, semolowaru, ditutup, etol, juanda, menanggal, arah, bikin, exit, krian, pln, selesai, wiyung, minggu, pagi, mengalir, mobil, gerbang, dgn, jaringan, pemadaman, taman, jembatan, nyala, pondok, coba, kartu, kawan, gang, lalin, info, turun, manukan, kasih, menit, raya, parah, perum, kawasan, pake, bandara, telkom, agung, gudang, kerahkan, pmk, royal, unit, wadungasri, bantu, rumah, kemarin, hujan, hati
Hidden	Trial-Error		Jumlah Neuron sebanyak 71
Output	1	2	Prediksi kelas negatif, prediksi kelas positif

Tabel 4.8. Model Neural Network di Tiap Neuron dalam 1 Hidden Layer

Neuron	Model
Neuron 1	$\hat{z}_1 = \frac{1}{1+e^{-(-0,075-1,007x_1+0,1416x_2-1,053x_3...+0,604x_{110})}}$
Neuron 2	$\hat{z}_2 = \frac{1}{1+e^{-(2,024-0,603x_1+0,708x_2-1,637x_3...+0,365x_{110})}}$
Neuron 3	$\hat{z}_3 = \frac{1}{1+e^{-(2,657+0,015x_1+1,048x_2-1,457x_3...-0,021x_{110})}}$
	⋮
Neuron 71	$\hat{z}_{71} = \frac{1}{1+e^{-(0,318+0,052x_1+0,136x_2-2,214x_3...+1,301x_{110})}}$

Tabel 4.9. Model Neural Network untuk Klasifikasi *Tweet* Sentimen Negatif dan Positif

No	Model
1	$\hat{g}_{\text{negatif}} = \frac{1}{1 + e^{-(-0,123 - 1,772z_1 + 0,12z_2 - 0,003z_3 \dots - 0,453z_{71})}}$
2	$\hat{g}_{\text{positif}} = \frac{1}{1 + e^{-(-2,008 + 0,027z_1 + 0,104z_2 + 0,54z_3 \dots + 0,530z_{71})}}$

Dengan x_1, x_2, \dots, x_{10} adalah variabel input berupa *term* yang sudah didapatkan melalui proses normalisasi (*preprocessing*). Neuron $z_1, z_2, z_3, \dots, z_{71}$ didapatkan dari perhitungan yang terjadi di dalam *hidden layer*. Adapun model fungsi aktivasi *sigmoid* sentimen negatif dan positif yang diperoleh ditampilkan pada Tabel 4.9. Model yang didapatkan dengan satu *hidden layer* dan 71 jumlah neuron kemudian digunakan untuk memprediksi data *testing*. Evaluasi kinerja model dari prediksi tersebut memberikan nilai kriteria pengukuran seperti yang ditampilkan pada Tabel 4.10.

Tabel 4.10. Ketepatan Klasifikasi Metode NN 1 Hidden Layer 71 Neuron

Akurasi	Presisi	Recall	Skor F1
80,00%	3,33%	80,40%	75,19%

Model yang didapatkan dengan satu *hidden layer* dan 71 jumlah neuron memberikan akurasi sebesar 80,00%. Sedangkan nilai presisi, nilai *recall*, dan nilai skor F1 nya berturut-turut ialah 3,33%, 80,40%, dan 75,19%. Menurut informasi yang telah disampaikan tersebut, dapat ditarik kesimpulan bahwa model klasifikasi neural network yang diperoleh bisa dikatakan dapat memberikan ketepatan klasifikasi yang akurat.

4.1.8 Perbandingan Metode Klasifikasi

Langkah terakhir yang dilakukan setelah mendapatkan hasil kinerja metode klasifikasi SVM dan Neural Network ialah membandingkan ketepatan klasifikasi. Tujuan daripada ini yaitu untuk menambah informasi tentang metode yang memberikan hasil klasifikasi terbaik untuk klasifikasi pendapat masyarakat terhadap pembangunan infrastruktur Kota Surabaya melalui Twitter. Berdasarkan

Subbab 4.1.6, diketahui bahwa untuk hasil terbaik metode SVM Non-linier menggunakan kernel RBF dengan parameter $C = 0,01$ dan $\gamma = 1000$. Adapun metode *Neural Network* menggunakan satu *hidden layer* dan jumlah neuron sebanyak 71 yang diperoleh dari pembahasan pada Subbab 4.1.7. Agar lebih jelas, perbandingan ketepatan klasifikasi metode SVM dan metode Neural Network ditabulasikan pada Tabel 4.11.

Tabel 4.11. Perbandingan Ketepatan Klasifikasi SVM dan NN

Metode	Akurasi	Presisi	Recall	Skor F1
SVM	92,67%	66,67%	92,24%	77,39%
NN	80,00%	3,33%	80,40%	75,19%

Menurut Tabel 4.11, dapat dilihat kinerja kedua metode klasifikasi dan dapat ditarik kesimpulan jika metode SVM Non-linier menggunakan kernel RBF dengan parameter $C = 0,01$ dan $\gamma = 1000$ memberikan kinerja yang jauh lebih baik. Hal ini berdasarkan nilai persentase akurasi, presisi, *recall*, dan skor F1 metode SVM Non-linier menggunakan kernel RBF dengan parameter $C = 0,01$ dan $\gamma = 1000$ yang lebih tinggi dibandingkan metode Neural Network satu hidden layer dengan jumlah 71 neuron.

4.2 Pembentukan Kluster

Penelitian pada subbab ini bertujuan untuk mengelompokkan data *tweet* berdasarkan *term* atau kata yang paling sering muncul di dataset disebabkan kata-kata tersebut sering dibicarakan oleh warga Surabaya. Secara umum, *term* yang paling sering muncul dalam 1500 data *tweet* ditampilkan pada Tabel 4.12.

Data yang berjumlah 1500 *tweet* dibedakan menjadi dua jenis, yaitu *tweet* bersentimen negatif dan *tweet* yang bersentimen positif. Algoritma yang digunakan untuk membentuk kluster yaitu algoritma K-Means. Pembedaan data *tweet* menjadi dua kategori ini didasarkan pada sentimen yang dimiliki tiap *tweet*. Data keseluruhan *tweet* ini mengandung dua sentimen yaitu negatif dan positif. Rasio *tweet* bersentimen negatif dan *tweet* bersentimen positif pada keseluruhan *tweet*

adalah 80%:20%. Maka data dibedakan menjadi 1200 *tweet* bersentimen negatif dan 300 *tweet* bersentimen positif berdasarkan rasio tersebut.

Tabel 4.12. Frekuensi Term

Term	Frekuensi	Term	Frekuensi
listrik	241	Mati	201
padam	181	Air	150
daerah	125	macet	138
mohon	116	perbaiki	77
masuk	51	Jam	114
kebakaran	76	Arah	145
tol	229	Pln	93
pdam	134	wiyung	55
jalan	173	Pagi	57
lampu	122	Nyala	63

Pada tahap pengerjaan, data sebelumnya di-*preprocess* dahulu layaknya Subbab 4.1.3. Kemudian, langkah selanjutnya adalah membuat *term document matrix* agar data siap diolah. Istilah dan struktur data *term document matrix* berbeda dengan struktur data *document term matrix*. Misalnya terdapat *corpus* dari dokumen-dokumen dan kamus kata yang mengandung semua kata yang muncul di dalam dokumen. Fungsi dari *term document matrix* ialah sebagai matrix dua dimensi yang barisnya merupakan kata-kata dan kolomnya adalah dokumennya. Pada data ini, yang dimaksud satu dokumen adalah satu *tweet*. Maka dengan begitu, setiap elemen (i, j) mewakili frekuensi dari term ke- i pada *tweet* ke- j . Pada setiap kelas akan dilakukan pengelompokan term menggunakan algoritma K-Means untuk mengetahui topik apa yang sering dibicarakan masyarakat Kota Surabaya.

4.2.1 Pembentukan Kluster Sentimen Negatif

Pengelompokan term pendapat masyarakat Kota Surabaya yang sudah melalui tahap *preprocessing* kemudian akan membentuk sejumlah kluster berdasarkan sentimen negatif dengan cara *trial-error*. Sebagai informasi, terdapat 2533 term dalam 1200 *tweet* diperoleh dari pembentukan *term document matrix*.

Merujuk pada Gambar 4.8 dan algoritma yang tercantum pada Subbab 2.9, dapat diketahui bahwa jumlah kluster optimum untuk *term* pendapat masyarakat Kota Surabaya sentimen negatif adalah sebanyak 2 kluster.

2. Metode Pseudo F

Selain mencari jumlah kluster optimum dengan metode Elbow, pencarian tersebut juga bisa menggunakan metode lain, salah satunya metode Pseudo F. Nilai-nilai Pseudo F berdasarkan jumlah klasternya ditampilkan pada Tabel 4.13. Apabila nilai suatu kluster dengan nilai kluster selanjutnya berbentuk siku tangan dalam grafik Elbow atau nilainya mengalami penurunan paling besar maka nilai kluster tersebut yang terbaik. Terlihat pada Tabel 4.13. bahwa nilai Pseudo F menurun drastis pada jumlah optimum dua kluster ke nilai Pseudo F dengan jumlah optimum tiga kluster. Oleh sebab itu, dipilih dua kluster sebagai jumlah kluster yang optimum.

Tabel 4.13. Nilai Pseudo F Jumlah Kluster Sentimen Negatif

Jumlah Optimum Kluster	Pseudo F	Jumlah Optimum Kluster	Pseudo F
2	46,2441	8	6,4733
3	23,0445	9	5,6447
4	15,3113	10	5,0003
5	11,4446	11	4,4848
6	9,1247	12	4,0629
7	7,5780	13	3,7114

Lima *term* anggota kluster yang berada paling dekat dengan *centroid* kluster pada kluster pertama dan kluster kedua adalah sebagai berikut:

klaster1: listrik, padam, daerah, mohon, pln.

klaster2: tol, macet, mati, air, pdam.

Dengan demikian, dapat diinterpretasikan bahwa sentimen negatif pendapat masyarakat Kota Surabaya ialah berpusat pada: listrik dari PLN sering padam, pintu keluar/masuk tol macet, dan air PDAM yang sering mati.

Pemadaman listrik sering terjadi dikala Bulan September, Oktober, sampai Januari. Hal ini dapat dikaitkan dengan pada kurun waktu tersebut iklim Indonesia memasuki musim penghujan. Hujan deras dan angin kencang kerap merusak kabel dan tiang listrik (Perdana, 2017).

Pintu keluar/masuk tol macet dikarenakan masyarakat menghadapi perubahan sistem pembayaran tol. Jasa Marga memutuskan bahwa mulai Oktober 2017, tidak ada lagi transaksi tunai (*cashless*) di gerbang-gerbang tol seluruh Indonesia. Semua transaksi di gerbang-gerbang jalan bebas hambatan (tol) akan menggunakan sistem elektronik (Faiz, 2017).

Aliran air PDAM pada Bulan September mengalami pengurangan intensitas laju atau bahkan mati pada tanggal-tanggal tertentu. Ari Bimo Sakti selaku Manager Sekretariat dan Humas PDAM Surya Sembada mengatakan, pada tanggal-tanggal tertentu akan dilakukan sejumlah pekerjaan, diantaranya penggantian travo existing 800 kVA menjadi 1600 kVA di IPAM Ngagel (Supingah, 2017).

4.2.2 Pembentukan Klaster Sentimen Positif

Pembentukan klaster sentimen positif ini mengalami proses yang persis sama dengan pembentukan klaster sentimen negatif. Pada *term document matrix* data sentimen positif, terdapat 879 term di dalam 300 *tweet*. Data teks sentimen positif ditampilkan secara visual menggunakan *wordcloud* seperti pada Gambar 4.8.



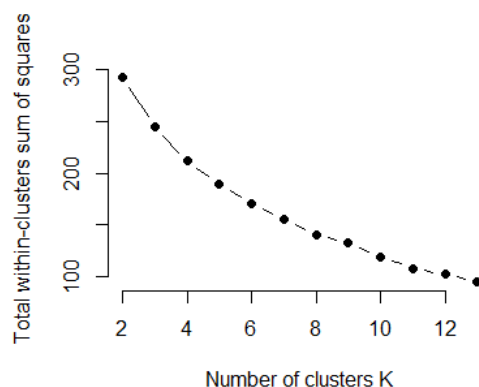
Gambar 4.9. Wordcloud Terms Sentimen Positif

Klaster yang dibentuk sebanyak 10 klaster yaitu seperti yang dilampirkan pada Lampiran 8. Tetapi, banyak topik yang berulang dan tumpang tindih. Agar

mendapatkan jumlah kluster yang optimum, selanjutnya diimplementasikan teknik pemilihan jumlah kluster optimum.

1. Metode Elbow

Berdasarkan algoritma metode Elbow pada Subbab 2.9, diperoleh plot Gambar 4.10.



Gambar 4.10. Elbow Kluster untuk Sentimen Positif

Merujuk pada Gambar 4.10, dapat diketahui bahwa jumlah kluster optimum untuk *term* pendapat masyarakat Kota Surabaya yang bersentimen positif adalah sebanyak 2 kluster.

2. Metode Pseudo F

Pencarian jumlah kluster optimum dengan menggunakan metode Pseudo F juga dilakukan pada data *term* yang bersentimen positif. Nilai-nilai Pseudo F berdasarkan jumlah klasternya ditampilkan pada Tabel 4.14. Apabila nilai suatu kluster dengan nilai kluster selanjutnya berbentuk siku tangan dalam grafik Elbow atau nilainya mengalami penurunan paling besar maka nilai kluster tersebut yang terbaik. Terlihat pada Tabel 4.14 bahwa nilai Pseudo F menurun drastis pada jumlah optimum dua kluster ke nilai Pseudo F dengan jumlah optimum tiga kluster. Oleh sebab itu, dipilih dua kluster sebagai jumlah kluster yang optimum.

Tabel 4.14. Nilai Pseudo F Jumlah Klaster Sentimen Positif

Jumlah Klaster Optimum	Pseudo F	Jumlah Klaster Optimum	Pseudo F
2	63,6659	8	8,9120
3	31,7261	9	7,7713
4	21,0795	10	6,8841
5	15,7562	11	6,1743
6	12,5623	12	5,5936
7	10,4329	13	5,1097

Hal sama dengan pembentukan klaster bersentimen negatif terjadi. Jumlah 2 klaster memiliki nilai Pseudo F paling besar maka 2 adalah jumlah optimum untuk klaster *term* bersentimen positif. Lima *term* anggota klaster yang berada paling dekat dengan titik pusat klaster pada klaster pertama dan klaster kedua adalah sebagai berikut:

Klaster1: jalan, perbaikan, arah, tol, pagi.

Klaster2: tol, lancar, arah, jembatan, info.

Interpretasi yang didapatkan adalah, bahwa masyarakat Kota Surabaya memberikan sentimen positif dan respon positif kala di pagi hari, jalanan, tol, dan jembatan lancar.

Balai Besar melakukan pembongkaran beton jalan, khusus kendaraan roda dua di Jembatan Tol Surabaya-Madura (Suramadu) dan akan melakukan perbaikan beton jalan yang udah dirusak di akhir bulan September (Indah, 2017).

Masyarakat antusias akan pernyataan dari Pemkot Surabaya yang menargetkan pengerjaan Jembatan Ratna, yang menghubungkan Jalan Ratna dengan Jalan Bengawan akan tuntas sebelum 10 November 2017. Ridlo Noor Wahab Kepala Bidang Peranangan dan Pengawasan Dinas Pekerjaan Umum, Bina Marga, dan Pematuan (PUBMP) Surabaya mengatakan, nanti di jembatan itu akan terpasang aksesoris berupa ornamen di bawah dan di atas jembatan supaya terlihat cantik (Indah, 2017).

BAB 5

KESIMPULAN DAN SARAN

5.1 KESIMPULAN

Berdasarkan analisis dan pembahasan yang telah dilakukan di bab sebelumnya, maka diperoleh kesimpulan sebagai berikut.

1. Hasil analisis sentimen terhadap data Twitter mengenai pembangunan infrastruktur Kota Surabaya dengan menggunakan metode SVM dan Neural Network.
 - a. Hasil ketepatan klasifikasi dengan parameter $C = 0,01$ dan $\gamma = 1.000$ mengatakan, persentase informasi yang diberikan oleh sistem terklasifikasi dengan akurat sebesar 92,67% tepat dengan informasi yang diharapkan peneliti. Nilai presisi 66,67%, nilai *recall* sebesar 92,24%, dan F1 sebesar 77,39%. Nilai evaluasi kinerja model ini dinyatakan bahwa term terklasifikasi dengan baik.
 - b. Model yang didapatkan dengan satu *hidden layer* dan 71 jumlah neuron memberikan kinerja evaluasi akurasi sebesar 80,00%, presisi sebesar 3,33%, sensitivitas sebesar 80,40%, dan skor F1 sebesar 75,19%. Dapat disimpulkan bahwa metode NN memiliki nilai evaluasi kinerja yang lebih rendah dari SVM, namun masih dapat mengklasifikasikan data dengan baik.
2. Sentimen negatif pendapat masyarakat Kota Surabaya ialah berpusat pada: listrik pln sering padam, pintu keluar/masuk tol macet, dan air PDAM yang sering mati. Sedangkan masyarakat Kota Surabaya memberikan sentimen positif dan respon positif kala di pagi hari, jalanan, tol, dan jembatan lancar atau telah dilakukan perbaikan.

5.2 SARAN

Saran yang diberikan atas hasil penelitian ini adalah:

1. Bagi penelitian selanjutnya, diharapkan melakukan *feature selection* setelah tahap *preprocessing*. Hal ini diharapkan agar *tweet* yang kata-katanya tidak baku atau salah ketik bisa diuraikan dan dijadikan kata baku. Tujuannya supaya semua informasi data diterima dan dapat diolah dengan baik, dan dapat meningkatkan tingkat akurasi klasifikasi.
2. Pengawas, Pemerhati dan Pemerintah Kota Surabaya dapat melakukan analisis sentimen masyarakat terhadap pembangunan infrastruktur menggunakan metode SVM karena metode tersebut terbukti memiliki tingkat akurasi yang baik.

3. Penelitian klasifikasi *text mining* yang ingin menggunakan metode *Neural Network* selanjutnya, sebaiknya menelusuri lebih dalam lagi kecocokan metode *Neural Network* terhadap *text mining*, seperti mencoba metode *Deep Learning*.
4. Penelitian *text clustering* selanjutnya sebaiknya membandingkan dua atau lebih metode klaster agar dapat membentuk klaster yang lebih akurat.

DAFTAR PUSTAKA

- Agustin, M., & Prahasto, T. (2012). Penggunaan Jaringan Syaraf Tiruan Backpropagation Untuk Seleksi Penerimaan Mahasiswa Baru Pada Jurusan Teknik Komputer Di Politeknik Negeri Sriwijaya. *Jurnal Sistem Informasi*, 2(2), 089-097.
- Auria, L., & Moro, R. (2008). *Support Vector Machine (SVM) as a Technique for Solvency Analysis*. Berlin: German Institute for Economic Research.
- Banerjee, A., Merugu, S., Dhillon, I. S., & Ghosh, J. (2005). Clustering With Bregman Divergences. *J. Mach. Learn Res*, 1705-1749.
- Basri, M. (2016). *Identifikasi Topik Informasi Publik Media Sosial di Kota Surabaya Berdasarkan Klasterisasi Teks Pada Twitter dengan Menggunakan Algoritma K-Means*. Surabaya: ITS.
- Bataineh, M., & Marler, T. (2017). Neural Network for Regression Problems with Reduced Training Sets. *Neural Networks: The Official Journal of the International Neural Network Society*, 1-9.
- Bekkar, M., Djemaa, H. K., & Alitouch, T. A. (2013). Evaluation Measure for Models Assesment over Imbalanced Data Sets. *Journal of Information Engineering and Applications*, 3, 27-38.
- Bholowalia, P., & Kumar, A. (2014). EBK-Means: A Clustering Techniques Based on Elbow Method and K-Means in WSN. *International Journal of Computer Application (0975-8887)*, IX(105), 17-24.
- Burges, L. (1998). A Tutuorial on Support Vector Machine for Pattern Recognition. *Data Mining and Knowledge Discovery*, 02, 955-974.
- Calinski, T., & Harabasz, J. (1974). A Dendrite Method For Cluster Analysis. *Communications in Statistics-Theory and Methods* 3, 1:27.
- Dhini, A., Hardaya, I., Surjandari, I., & Hardono. (2017). Clustering and Visualization of Community Complaints and Proposals using Text Mining and Geographic Information System. *3rd International Conference on Science in Information Technology (ICSITech)*, 132-137.
- Dumais, S. (2004). Latent Semantic Analysis. *Annual Review of Information Science and Technology*, 188-230.

- Ezunumah, B. (2013). College Students' Use of Social Media: Site Preferences, Uses and Gratifications Theory Revisited. *International Journal of Business and Social Science*, 28-29.
- Faiz. (2017, September 08). *Atasi Macet di Tol, Transaksi Non Tunai Belum Solusi Akhir*. Diambil kembali dari Suara Surabaya: <http://www.suarasurabaya.net/mobile/kelanakota/detail/2017/193105-Atasi-Macet-di-Tol,-Transaksi-Non-Tunai-Belum-Solusi-Akhir>
- Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advances Approaches in Analyzing Unstructured Data*. United States of America: Cambridge University Press.
- Fithriasari, K., Iriawan, N., Ulama, B., & Sutikno. (2013). On the Multivariate Time Series Rainfall Modeling Using Time Delay Neural Network. *International Journal of Applied Mathematics and Statistics*.
- Gartner. (2017, September 12). *Definition: What is Big Data?* Diambil kembali dari Definition: What is Big Data?: <https://research.gartner.com/definition-what-is-big-data?resId=3002918&srcId=1-8163325102>
- Gunn, S. (1998). *Support Vector Machines for Classification and Regression*. England: ISIS.
- Habibi, R. (2016, April 1). Analisis Sentimen Pada Twitter Mahasiswa Menggunakan Metode Backpropagation. *INFORMATIKA*.
- Haerdle, W., Prastyo, D., & Hafner, C. (2014). Support Vector Machines with Evolutionary Feature Selection for Default Prediction. Dalam *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics* (hal. 346-373). United States of America: Oxford University Press.
- He, W., Wu, H., Yan, G., Akula, V., & Shen, J. (2015). A novel social media competitive analytics framework with sentiment benchmarks. *Information & Management*, 12.
- Hemalatha, I. (2012, July-August). Preprocessing the Informal Text for Efficient Sentiment Analysis. *International Journal of Emerging Trends & Technology in Computer Science (IJETICS)*, 1(2), 58-61.
- Hsu, C. W., & Lin, C. J. (2002). A Comparison of Methods for Multiclass Support Vector Classification. *IEEE Trans. Neural Network*, 13, 415-425.

- Indah, R. (2017, September 2014). *BBPJN: Perbaikan Jalan Beton Gunakan Metode Mutakhir Khusus*. Diambil kembali dari Kelana Kota Suara Surabaya FM 100: <http://kelanakota.suarasurabaya.net/news/2017/193795-BBPJN-:-Perbaikan-Jalan-Beton-Gunakan-Metode-Mutakhir-Khusus>
- Indah, R. (2017, Oktober 18). *Target Pengerjaan Jembatan Ratna Tuntas Sebelum Hari Pahlawan*. Diambil kembali dari Kelana Kota Suara Surabaya FM 100: <http://kelanakota.suarasurabaya.net/news/2017/194753-Target-Pengerjaan-Jembatan-Ratna-Tuntas-Sebelum-Hari-Pahlawan>
- Kristalina, P. (2015). *An Introduction to Smart Cities (Pengenalan tentang Smart City)*. Surabaya: Politeknik Elektronika Negeri Surabaya.
- Kumar, A., & Paul, A. (2016). *Mastering Text Mining with R*. Birmingham, Mumbai: Packt Publishing.
- Kurniawan, T. (2017). *Implementasi Teks Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Media Mainstream Menggunakan Naive Bayes Classifier dan Support Vector Machine*. Surabaya: ITS.
- Kwartler, T. (2017). *Text Mining in Practice With R*. New Jersey: John Wiley & Sons Ltd.
- Lauder, T. K., & Dumais, S. T. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of The Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, 211-140.
- Lebart, L. (1998, December). Text Mining in different languages. *14*(4), 323-334.
- Lidya, S. K., Sitompul, O. S., & Efendi, S. (2015). Sentiment Analysis Pada Teks Bahasa Indonesia Menggunakan Support Vector Machine (SVM) dan K-Nearest Neighbor (K-NN). *Seminar Nasional Teknologi Informasi dan Komunikasi 2015 (SENTIKA 2015)*.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- Liu, B. (2015). *Sentiment Analysis Mining Opinions, Sentiments, and Emotions*. United State of America: Cambridge University Press.
- Madhulatha, T. S. (2012). An Overview on Cluster Methods. *Jurnal of Engeenering*, 719-725.

- Mandal, S., & Banerjee, I. (2015). Cancer Classification Using Neural Network. *International Journal of Emerging Engineering Research and Technology*, 172-178.
- Manning, C. D., & Raghavan, P. (2008). *Introduction to Information Retrieval*. Stuttgart: Cambridge University Press.
- Mercer, C., & Li, C. (1909). Foundations of Positive and Negative Type, and Their Connection With The Theory of Integral Equations. *Philosophical Transactions of the Royal Society of London*, 25, 3-23.
- Mergel, I. (2013). A Framework for Interpreting Social Media Interactions in the Public Sector. *Government Information Quarterly*, 327-334.
- Monarizqa, N., Nugroho, L. E., & Hantono, B. S. (2014, October). Penerapan Analisis Sentimen Pada Twitter Berbahasa Indonesia Sebagai Pemberi Rating. *Jurnal Penelitian Teknik Elektro dan Teknik Informasi*, 01.
- Morissette, L., & Chartier, S. (2013). The K-Means Clustering Technique: General Consideration and Implementation in Mathematica. *Tutorials in Quantitative Methods for Psychology*, 12-24.
- Oktara, D. (2017, Mei 19). *TEMPO.CO*. Dipetik July 15, 2017, dari BETA TEMPO.CO:
<https://nasional.tempo.co/read/news/2017/05/19/058876933/menko-darmin-sebut-surabaya-contoh-smart-city-di-indonesia>
- Orpin, A. R., & Kostylev, V. E. (2006). Toward a Statistically Valid Method of Textural Sea Floor Characterization of Benthic Habitats. *Marine Geology*, 209-222.
- Pang, B., Lillian, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. *Empirical Methods in Natural Language Processing (EMNLP)* (hal. 79-86). Philadelphia: Association for Computational Linguistics.
- Perdana, D. (2017, Oktober 22). *Angin Kencang di Driyorejo Rusak Bangunan dan Kabel Listrik*. Diambil kembali dari Kelana Kota Suara Surabaya:
<http://kelanakota.suarasurabaya.net/news/2017/194904-Angin-Kencang-di-Driyorejo-Rusak-Bangunan-dan-Kabel-Listrik>
- Prasetyo, S. E. (Penyunt.). (2017, January 11). *JawaPos*. Dipetik September 13, 2017, dari www.jawapos.com:
<https://www.jawapos.com/read/2017/01/11/101467/pemkot-surabaya-geber-proyek-infrastruktur-pengembang-terbantu-akses-yang-lebar>

- Putranti, N. D., & Winarko, E. (2014). Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine. *IJCCS*, 91-100.
- Rismaharini, T. (2013). Pengembangan Infrastruktur Kota Surabaya, Antara Problem dan Solusi. *Seminar Nasional Aplikasi Teknologi Prasarana Wilayah 2013* (hal. 17-27). Surabaya: Program Diploma Teknik Sipil Fakultas Teknik Sipil dan Perencanaan Institut Teknologi Sepuluh Nopember.
- Robandi, I., & Gusti, R. A. (2008). *Peramalan Beban Jangka Pendek untuk Hari-Hari Libur dengan Metode Support Vector Machine*. Surabaya: ITS.
- Rumiarti, C. D. (2017). Segmentasi Pelanggan Pada Customer Relationship Management di Perusahaan Ritel : Studi Kasus PT. Gramedia Asri Media. *Journal of Information Systems*, 1-10.
- Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis, Teori dan Aplikasi*. Graha Ilmu.
- Saraswati, N. S. (2013). Naïve Bayes Classifier dan Support Vector Machines Untuk Sentiment Analysis. *Seminar Nasional Indormasi Indonesia. SESINDO*.
- Sarkar, D. (2016). *Text Analytics With Python*. Bangalore, Karnataka: Apress.
- Scholkopf, B., & Smola, A. (2002). *Learning With Kernel : Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, 427-437.
- Sun, Y., Kamel, M. S., & Wang, Y. (2006). Boosting for Learning Multiple Classes with Im-balanced Class Distribution. *Sixth International Conference on Data Mining (ICDM'06)*, (hal. 421-431).
- Supingah, I. (2017, September 22). *Siap-siap Gangguan Air PDAM Surabaya pada 22-23 September*. Diambil kembali dari Kelana Kota: <http://kelanakota.suarasurabaya.net/news/2017/193734-Siap-siap-%E2%80%8EGangguan-Air-PDAM-Surabaya-pada-22-23-September>
- Surabaya, W. R. (2017, Maret 29). *surabaya.go.id*. Diambil kembali dari [surabaya.go.id: http://www.surabaya.go.id/berita/32584-musrebang-](http://www.surabaya.go.id/berita/32584-musrebang-surabaya.go.id)

surabaya-2017-fokus-percepatan-pembangunan-infrastruktur-berwawasan-lingkungan-

Susanto, H. (2014). *Visualisasi Data Teks Twitter Berbasis Bahasa Indonesia Menggunakan Teknik Pengklasteran*. Surabaya: ITS.

Widhianingsih, T. (2016). *Aplikasi Text Mining untuk Otomatisasi Klasifikasi Artikel Dalam Majalah Online Wanita Menggunakan Naive Bayes Classifier (NBC) dan Artificial Neural Network (ANN)*. Surabaya: Institut Teknologi Sepuluh Nopember.

Wikipedia. (2017). Diambil kembali dari Wikipedia Ensiklopedia bebas: <https://id.wikipedia.org/wiki/Twitter>

LAMPIRAN

Lampiran 1. Istilah total sebanyak 2.824 istilah pada 1500 *tweet*

acara, bazar, gimana, jagir, kampung, listrik, nihpln, padam, sidomukti, aneh, enek, kota, metropolitan, parkir, daerah, koq, sidokepung, slalu, yakoq, butuh, internet, kompln, mohon, pekerjaan, perhatiannya, telpon, assalamualaikum, conek, telp, box, culvert, danau, dianggap, ekor, kemacetan, menganti, penyebabnya, smpai, unesapembangunan, atum, diganti, gant, jgn, kolam, malas, masuk, pasar, pengelola, renang, tiket, tulisannya, giat, kabel, kebakaran, olah, suramadu, tkp, tol, hijaunya, kayaknya, lama, macetlampu, padat, pendekmerahnya, putroagung, ala, berbahaya, jemursari, kadarnya, pdam, pemasangan, pengamanan, penind, pipa, proyek, santika, bocor, galian, gara, iya, jalan, kali, terulang, eduaubdedubua, pantauan, rungkut, selatan, terkait, wonorejo, cctv, iskandar, karang, muda, nyamplungan, pelanggaran, perempatan, sidorame, sultan, tembok, adil, dagang, dipakai, eduaubdedubudjl, jalur, pedestrian, pickupnya, hari, kutisari, malam, pengaturan, tiang, data, dipercy, promo, propert, public, sekian, sms, sulit, terima, berhenti, kecuali, konsultasi, kpp, pelayanan, penjelasan, pratama, total, wonocolo, lampu, mati, pring, tambak, timur, beralih, biar, feeder, masyarakat, nunggu, pemkot, realisasikan, trem, trunk, air, airnya, barat, baskara, dibantu, got, item, mtnwn, spt, bentar, lagilampunya, macet, potensi, sari, tanjung, baw, dibutuhkan, industri, satelit, arus, lintas, perbaikan, dicopot, merr, mmg, penertiban, pisan, rambunya, rkt, ruwetruwet, tindakan, hasil, persebaya, gubeng, sumuk, djuanda, elshintanewsandtalk, sonoranetwork, toll, dupak, lancar, ramai, babatan, cnt, culvertpdhl, depo, dijadikan, indah, kesalahn, motormobil, perumahan, pnutupan, ribuan, terfataladl, full, jlrandu, kenjeran operator, pengaduan, pulsa, sdtopo, separuh, terpotong, wetan, detlen, iki, jam, min, ini, jllakarsantri, penutupan, wil, hidupnya, jaya, ngagel, etoll, gate, htt, menerima, pariwisata, payah, pembayaran, pintu, tunai, antrean, bola, bolasaran, card, hrsnya, jemput, kelemahan, pakai, petugas, readernya, sistem, buang, gedangan, kantor, sadar, sembaranganpemandangan, smpah, mulyosari, nohp, nopel, prima, respon, seminggu, bersih, dikuras, semolowaru, wilayah, klu, repot, tolgt, bbm, dadan, desember, harga, jenis, memutuskan minyak, pemerintah, solar, ssinfo, tanah, ditutup, etol, etollgate, gunanya, juanda, menanggal, nemene, operasioalahkok, capek, dicheck, diperbaiki, ditindak, doang, kerusakan, lanjuti, sbg, untk, warga, arah, bikin, dibedakan, exit, gresik, krian, muter, saja, plus, bon, bukti, claim, kehabisan, kertas, org, pngelola, print, maksimal, odprt, oktober, pln, selesai, transmisi, diperlebar, jalannya, menyumber, nginden, yani, aliran, ditangani, gangguan, wiyung, bougenville, compalin, mengajukan, minggu, ngeyel, pagi,

sbtu, tropodo, wisma, darmo, ingo, mengalir, permai, simpang, asa, derek, kecelakaan, kepadatan, mobil, dipush, gini, mahal, mahalcobalah, menguntungkan, provider, quota, rakyate, sekali, tarif, dievaluasi, harysnya, hrus, konfirmasi, lah, membuang, perlumohon, efek, macetnya, separah, dipaksakan, gerbang, kurang, merah, napsu, tenaga, dgn, kuecil, sekitaran, baru, mastol, wajar, jadwal, jaringan, listrikblm, pemadaman, perawatan, rewwin, tau, nih, semuanyaaaa, sidosermo, dipo, sidotopo, stasiun, disaat, libur, matinya, shubuh, bgs, duduk, iupan, jln, ksh, kykny, organisir, panas, perahunya, rompi, safety, suroboyo, taman, tempa, wisata, hijrah, transportasi, umu, fasilitas, menyusahkan, milik, orang, pinggir, pribadi, setuju prinsipnya, jembatan, berangkat, delay, flight, lion, pswt, subbdj, maritim, nyala, padamsejak, pondok, tadisampa, coba, kamera, mendukung, menguji, program, resolusi, tilang, lain, migrasi, pekan, pelanggan, perak, sinyal, utara, gempol, penerapan, disusul, gerimis, bayar, cash, diambil, emoney, kartu, kuran, mbuka, nempel, nempelnya, portal, call, center, kraton, maaf, menghubungi, sibuk, terussss, bergelombang, ess, jlannya, scorpioenjoy, trosobo, mbaj, sehat, terbentuk, kenaikan, sumo, kawan, kembang, kuning, matidi, alamat, bos, gang, gangguan, undaan, mantabalhamdulillah, mas, pengaspalan, bypass, kriyan, lalin, menjelang, desa, info, lalu, membawahi, sarirogo, sukodono, aja, dilebak, gbr, lemotsdh, restart, semuasangat, telkomsel, terbaiktermahal, berjualan, mastrip, mesin, premium, rusak, spbu, angka, sensi, ditarik, uang, wewenangnyanya, wilgresik, wilsby, blkg, jalanan, nggak, pikirkan, rsud, spy, suwandhie, bbpjn, beton, khusus, metode, mutakhir, odppr, benowo, kandang, truss, menutup, mutiara, ngopak, perlintasan, polsuska, turun, dlm, rangka, rekayasa, memakai, maksudnya, sawotratap, semalam, sih, kendaraan, luberan, menggenang, pengendara, polisi, pos, terg, tikungan, alhamdulillah, wes, pula, senin, kec, kemaren, manukan, menetes, ranu, tandes, basra, belok, bungkul, istimewa cenderung, kbs, komb, ksapi, merambat, sepi, sumoharjo, urip, dilangsungkan, kegiatan, kerjasama, bantuanx, kasih, trima, kel, kenapa, keputih, kabar, menit, sambikerep, surabaya, antri, blum, ditambahantri, guebeng, kai, kepanggiloket, lokalyg, loket, pesen, ribuanhari, kedam, kedameanlegundi, pnjang, polsek, raya, sisi, antrian, dibuka, eduaubdedubu, alfamart, area, habis, mandiri, cctvto, kelemahannyakrn, minjemi, msh, senang, takut, uda, infopl, auri, informasi, semambung, terimakasih, parah, romokalisari, foto, magersari, pekat, penampakan, perum, ngadih, niat, bratang, kawasan, sekitarnya, wow, buru, kecepatan, kurangi, nambah, pake, siap, timer, udh, ush, bandara, jumat, kayak, skrang, berfungsi, kertajayadharmawangsa, pada, penyebab, betulkan, dikawasan, korban, mksh, roboh, telkom, display, dkt, jalanrusak, kenyataan, lokasi, park, penuh, ptg, tertulis, tpt, tvri, wildmansensor, zone, agung, gudang, kerahkan, pmk, royal, unit, wadungasri, diberangkatkan, dikasih, emosi, roti, ruang, sesak, tunggu, walo, aloha, bunderan cenderung, waru, baru, kalirungkutpanjangjiwo, kedung, kurleb, lubang, meter, pas, selebar, infonya,

kupang, share, simogunungsimolangitdk, asas, bahagiadat, bergotongroyong, budaya, istiadat, tercipta, dalam, motor, sepeda, terbakar, update, penuhhhhh, hidup, ajamohon, barulah, demo, era, instansi, menuruti, pendemo, risma, sich, byk, donk, menyelwsaikan, solusi, gemblongan, karamenjanganggapadahal, pnumpangnyabahaya, alohawaru, bermasalahkluar, kecilbisa, mandi, skala, bantu, dukuh, forward, sebagai, tlg, agungjln, ayani, gayungan, klok, kndaraan, lebar, mngalah, mnuju, msih, msjid, prkir, sisipan, balong, panggang, sekitarnyaada, dan, dari, jemundo, sambisari, sawunggaling, denza, elektronik, mengabarkan, perdana, reporter, suara, bapak, cepat, memperbaiki, merespon, pandugo, trafo, dicek, hmpir, karangrejo, tlong, voltase, wnokromosudah, eotkaf, komplain, menteri, peralatan, karangmenjangan, kasihan, kebijakan, psr, tentu, tepatnya, tpi, bauh, blom, dtambah, krisis, mmprihatinkn, mnyengat, pakal, sampah, sep, sjak, sngt, sumberrejo, tgl, tpa, wrga, pemadamannya, brizzi, rembang, tolponya, akibat, internasional, ngurah, penumpang, rai, terdampak, kecilamp, king, mingguansampai, residence, rumah, safira, sperti, barang, cacat, diadakan, gea, lelang, bantuannya, bendul, menyala, merisi, padamsmoga, terkaitterima, yah, dri, makan, surabayamalang, jaln, pembuatan, pengkajian, trotoar, tunjungan, ulang, dsk, gede, bur, kebanyakan, soekarno, banteng, bulak, tenggumung, upayakan, dsktnya, kena, pengerjaan, prapenbackhoe, prapentenggilis, rabu, lebaniwaras, msk, wringinanom, hubungi, santunanjasaraharja, asri, candra, wadung, berja, kualitas, pasang, pelayananmubisa, perbaikan, surya, tersier, thn, banget, berbek, lancaaar, bnyk, boxculvert, lho, lihatnya, perhatian, sumpek, kendaraanberhati, merayapkarena, volume, ahmad, kereta, rel, rsi, badai, irma, juta, odprs, putus, aktif, beli, bri, gempa, mggu, tnyt, berkat, bravo, etilang, kepolisian, lanjutkan, met, siangtertib, adlh, akses, legundi, apapun, bisa, kluargakerjaan, bendungan, clear, completed, day, ruas, menyalathanks, seputar, jlnkepala, pecerahan, tadimohon, wage, ambles, mengenaskan, merana, nasib, aspalnya, bekas, boulevard, kertajaya, samping, ayuk, budayakan, garis, penyebrang, utama, antre, kend, sabar, comberan, dharmahusada, layak, pakai knp, naik, putuskan, jammalam, jamsiang, jawabnya, lampubagaimana, margorejo, masi, tanggung, bagaimana, cimanuk, kerja, pulang, dps, angkot, cilik, jatisrono, jembatane, alam, citraland, galaxy, kerep, sambu, gerangan, diizini, parahlho, pemagku, satunya, sukolilo, bunder, jalurmacet, kebomas, lahan, lantas, pny, suwandi, juga, plaza, jlnkenjeran, makam, jati, pemberitahuantx, bln, dijln, garudatp, ktr, layanan, perkembangan, troublelbih, bergerak, cuacanya, liat, ploso, benar, peka, tutup, dipake, simpati, dilengkapi, kaki, pejalan, wadah, giliran, donkini, tenggilis, biru, tenda, pasuruan, rejoso, timbang, diinfokan, ijin, lapor, pilang, error, gandungadi, jaringannya, kasi, kemarin, nomer, penga, telkomindonesia, blgnya, beban, berjalan, ketutup, lbh, meleak, puluhan, rupiah, seeh, blok, its, adakan, cfd, disepanjang, rapi, tertata, alternatif, giri, lwt, mayjen, pelabuhan, tembus, veteran, bangil, anggawangi,

dusun, kab, kweni, yacc, answer, bedakan, bertamunorma, etika, harta, kepemilikan, kudu, masy, pripun, regulasi, sosial, suggest, kipaskipas, moja, panaspanas, gbtsupporter, infrastuktur, pembangunan, stadion, bahaya, geriwis, gitu, gunungsari, roda, digiatkan, ektp, geje, online, registrasi, sim, ujungnya, blega, aman, dmna, flazz, subuh, mengecil, normal, sebulan, tiba, greges, imbas, hadehhhh, sampek, saur, semoga, tugas, brgkt, dibandara, juandashrsnya, skrng, smpe, tanda, berbekimbas, hrs, putar, terlanjur, bea, bsk, datangi, dganti, dgantiapa, ganti, katax, meteran, muka, ptgs, jual, alat, berat, graha, karna, sampurna, biarpun, deket, dikitlah, hijau, tetep, waspada, bukit, lidah, nya, anak, begini, jammenyusahkan, jelidro, kecilhampir, trmksh, lepas, melambat, penyempita, sahabat, terminalril, didaerah, response, dipotong, ditambah, inet, sebelumnya, tagihan, prigen, semalem, skr, noor, sedih, syamsudin, tertinggal, antriannya, bilka, ditinjau, ngangel, peran, stopan, dmmp, rejo, burukkerusakan, diluar, karangasem, losssdh, mnc, respondaerah, rmh, dpan, sampoerna, bca, cari, gto, mundur, dikenakan, pengguna, seksi, sosialisasi, uji, jummeduaubdedubuad, lahseng, lak, mesti, progran, rakyateajur, bgt, langsung, mbalesnya, time, hujan, kondisi, licin, pelan, diseluruh, mengalami, selamat, temanteman, dsuarakan, krena, numpuk, suka, tolong, trend, rosan, pemberitahuan, sawotratapedangansidoarjo, sktr, tdl, saingan, teruscoba, bandarejo, dianya, lupa, segini, masnunggu, nyamukmelekan, serangan, ditunggu, iam, kak, mochismail, salam, ylkiid, kakek, ngalir, cuman, eee, lha, pepelegiwaru, sebentar, pkm, ampun, susahnya, makasih, pak, ambrol, gantian, lajur, macettt, tinggal, heavy, most, mukti, blimbing, flyover, lawang, siwalankerto, bantuan, bermasalah, menindaklanjuti, ruko, squ, tidar, biarp, entry, kdg, pengambilan, tdk, beroperasi, kertosono, gerah, listriknga, mlm, mnta, sememi, tdur, atau, kfc, lift, pembawa, penyeberangan, susah, tanggapanya, bunderrr, tolll, jadwalnya, subjkt, culvertdirapikan, jalanlalu, manajemen, pdhl, pmbgunanny, psang, salahmestinyadibongkar, shg, babon, bumpy, hati, maklum, rute, truk, geluran, kmrn, sore, bentang, bertindak, gelap, gulita, kejahatan, lampunya, buka, menyebabkan, panjang, sitem, wakt, porong, dinas, karaoke, nav, patmosusastro, soetom, merisiampjagir, aturan, kinerja, membantu, tambahan, asih, babat, depot, lamongan, licinkemacetan, soto, bnget, bntu, kcil, kesusahan, kmi, matiinyala, kebun, sarangan, strowbery, terpantau, biaya, dikirim, home, indi, kecewa, sesuai, sgt, aratampmatinya, kra, pengerukan, probolinggo, brigjen, eduaubdedubuab, giant, siapsiap, katamso, bgm, drop, off, pondokcandra, rewwinrungkutwadungasrisabar, yaa, pertggjwb, tersendat, infoplninfo, jend, joyoboyo, medaeng, sudirman, munguti, senayan, baik, eskalasi, pelaporan, terpenuhi, indonesia, layananprimejasaraharja, asap, hitam, tebal, enam, getra, sibu, sblmnya, lumayan, merayap, tunaihelp, bangunan, dibiarkan, harmoni, ikutan, nutup, selokan, keputransangat, membahayakan, jasa, lurus, marga, marka, putusnya, salahmarka, andai, krcis, lucunya, mbil, messi, msuk, mtr,

percaya, pke, sbycukup, aspal, bendo, ketintang, pukul, seger, wib, romo, patmosusastrolokasi, tpatnya, pasir, sidoar, skitar, wedi, pengecekan, pengecekanmati, pompa, ada, ndk, pam, wonokusumo, saldo, dibilang, digempol, etollkatanya, kartunya, cokroaminoto, konstruksi, kasian, sekolah, toiletnya, dampak, debit, wiyng, padamkondisi, penanganannya, diphpin, doankwhy, dtgcm, mintp, smpk, speedytlp, akp, evakuasi, kanit, keperdulian, muatan, oleng, shukiyanto, truck, beres, dipercepat, miris, order, perbaikanya, prapen, nota, salahsatu, chandra, yang, harap, rukem, beberpa, bersepeda, mencoba, situasi, cuci, melih, djlkagung, gan, kos, suprpto, tetespun, trpaksa, viiie, bertanggung, jawab, karung, timbun, dinyalakan, kembalitrim, knapa, arteri, tumben, asuh, dala, ribud, tertera, token, mendrip, permata, kira, ciamik, nopolnya, soro, diatasi, kedurusharap, pol, apa, saurada, bagaiman, banjarsawah, siang, tegalsiwalanprobolinggo, dilebarin, disuruh, trotoarnya, bau, berasap, kerot, majapahit, mojokerto, totok, halo, hilang, kapasitas, sekarang, signal, bebas, hambatan, rembanggempol, asalan, kualitasnya, terkesan gowes, beru, dihilangin, kuatir, mbak, mendadak, ngerem, bubuk, busuk, deogone, limbahbel, pembasmi, solusinya, tuhtabur, gayung, infonyaterima, amit, bklsby, detik, dikec, dperbaiki, iniinilah, jmbtn, mnju, panjangnya, sbelm, infokan, penu, laporan, pusat, tanggapan, sumengko, thanks, baby, bumi, galaxi, sakit, skitarnya, tibatiba, tindak, balita, denda, lansia, telat, telingackckckck, tetangga, dgantiap, kembali, listriknya, posisi, medokan, pelebaran, upn, penanganan, purimas, bahu, tangki, bbrp, gatetol, kejapanan, tempel, terbuka, eduaubdedubuad, count, dihilangkan, down, cuaca, pembatas, perbaikn, berkeliling, mengakomodir, nusan, duty, equ, massal, penggunaan, penting, publik, shgg, kelihatanny, tergesa, cerah, dek, diikat, kapal, keselamatan, ketapang, penyebrangan, blkngan, ditgh, kcl, kumat, lgr, mampet, oasis, palm, jojoran, dam, disebabkan, dong, mayor, dibagi, kalteng, duh, eduaubdedubuegengges, loh, pakis, amp, bawa, bingung, merepotkan, non, tersedia, totalitas, alfa, menjuntai, midi, city, didukung, disempurnakan, eduaubdedubud, kekurangan, smart, tenan, terlepas, dibenerinnya, jemur, mengakibat, sekring, suwun, agustus, alasan, benowoyg, cover, kerjakan, kedurus, smp, toserba, weta, diapaapakanmbok, ditutupmacet, ditutuppdl, jlunny, keluar, pahamnant, menitan, perumtas, wonoayu, berlanjut, deheduaubdedubuadeduaubdedubuad, proyeknya, gadung, hindari, jlrandegansari, yaaaa, banyu, gerak, jiwo, lancarrgak, brpa, diteruskan, macetpintu, urippiye, dijanjikan, matibikin, telponnya, lebihada, sejam, kemahalan, mlh, ntar, rugi, tollkalau, sumur, sidorejo, alternatifspy, curvert, disanajln, dulubox, hrsny, mangkrak, wiyungmenga, kanan, median, pekerja, diphpi, berikut, september, tanggal, wilayahwilayah, mengecewakan, pelayananmu, sungguh, wahyudi, gembili, melaporkan, sbymhn, kejawan, putih, tambakkecmulyorejo, jembayan, anom, dsn, joho, keboan, kaya, macett, mgu, terakhirtol, dilindungi kabeltidak, umur, infrastruktur, kebutuhan, manusia,

menciptakan, pokok, sehingg, diperbaharui, dns, keama, kpd, krglo, krgploso, smg, trm, merugikan, baruselalu, dijam, istirahat, mulyorejo, tegal, bgitunahklo, blg, etollapa, etollkok, invalidkata, mace, nyoba, petugasnya, porongpandaan, lobangmungkin, pacar, forklift, melintas, stikom, pkl, sambisarikellontar, men, kapan, slesai, garasitapi, hajat, mengganggu, ngadain, nikahan, pengend, datang, mall, sungai, bank, penyedia, nyebrang, semua, gendam, pikir, wajib, ditata, ibuheranada, infrastrukturnya, pikirnya, terbalik, penerangan, sedenganmijenkrian, tanggulkec, inimohon, kendala, perhatiannyaada, morowudimacetsaya, terkena, indihome, telkomteknisi, cuvlet, muat, sebelah, loc, pejalanan, pertamanan, trim, dimana, gangguannyahr, plnkasian, pra, kesini, pembersihan, upaya, berlomba, countdown, disalahgunakan, duluan, start, angkat, kagak, gading, disambut, lgs, warujuanda, bangkalanapkh, jemb, kmacetan, stuasi, teruraizuwuun, cerita, darurat, sbgmn, jangkrik, mie, peng, sungkono, kalianak, makasi, updatenya, jrngang, pembrthuan, pmdman, pmliharaan, prng, rayon, yaaa, berada, insfratraktur, menata, didepan, bossternyata, trims, ditimbulkan, hukum, hukumannya, kemacetanjalan, kurangakibat, tonase, sby, kendaraan, berlumpur, berwenang, jatuh, ujun, kusbang, nopol, tersangkut, tindaklah, property, emang, lelet, neh, dar, bantuanya, cms, gtomohon, infonyathxeduabdedubuf, okt, prabayar, enak, sak, sma, trimurti, udele, adeeemmmm, hjn, lncr, mendung, prapatkurung, ringan, banharsugihan, stop, citraraya, jlmayjen, loop, unesa, yonosuwoyodepan, hadiah, kejutanterima, pakuwon, ampel, disisakan, frontage, kampus, road, sunan, toppp, uin, ajaib, hahahah, kan, keciil, kingsafira, lagi, sebulan, dah, fiuh, hiks, untunglah, dibawah, etolltp, ormalisasi, proses, candipuro, jarit, lumajangjawa, pasirian, uranggantung, arusnya, darmokali, girder, pengalihan, ratna, sabtu, eduaubeedubua, enaknya, keramiknya, late, post, tambaksari, jarak, pandang, bfunksi, memudahkan, perbatasan, silahkan, tnnal, jas, brarti, menyalahgunakan, telepon, teleponnya, dikota, lihat, imbunnya, skg, clogging, dibwh, tujuan, nyuci, action, bnyak, mendekati, mestinya, bal, blgkartu, ptugas, tjd, maps, nama, landai, tembaan, channel, lokal, barusan, keterangan, senkom, buruknya, contoh, dibongkar, dimundurkan, perencanaan, setahun, fotoin, gorong, hehe, krna, sayang, sidak, tercinta, walikota, disini, jawa, polusi, lagisampe, ngantuk, nunggunya, deras, jalaan, jembtan, bocorpas, bronggalan, genangan, indomaret, kyaknya, sdkit, tngah, memenuhi, mengantiwiyungsemestuny, admin, konfirmasinya, sept, tlpn, connect, daritadi, kalanganyar, seda, wifi, lampuknp, perumh, kiri, sbilm, hany, nyaman, tertib, thenks, lebak, didalam, khawatir, terkini, mmpung, musim, psti, terganggu, keruh, pdak, perusahaan, singkatannya, bus, dilewati, kabuh, andong, gatenya, terminal, pengecoran, pertigaan, randegansari, tlapak, janji, balongsari, bibis, buntaran, culver, pngerjaan, stiap, trsambung, hri, jugruk, rejosari, trhitung, ckckckckck, diginiin, genap, bedapintu, hbs, hrg, ktnya, promonya, rbpintu, security, denger, lengang, milih, pahlawan,

panggung, perjalanan, rajawali, abis, dibuang, jaman, kuota, lgkan, mreka, murahny, ngejar, now, simcard, jadi, mengantiwiyung, plebaranpmbgunan, terencana, hujankampusbunair, mulaimati, ngetes, padamada, hape, hapenya, nebeng, bertahap, surabayagempol, akbar, masjid, bundaran, eduaubdedubuaeduaubdedubue, mengadu, speedy, stl, teknisi, baruhelp, nich, hrpn, hubgangguan, infrmsi, masak, mndi, rngkt, citilink, halim, kemungkinan, penumpukan, perdanakusuma, bersabar, eduaubdedubuaeduaubdedubua, mayjend, muhammad, hem, mengarah, tolnya, uantri, madiun, tarifnya, beginisaya, keberatanroyal, tarik, blank, spot, supermall, titik, miring, bla, cashless, daftar, dipersulit, ditolak, dukung, eduaubeedubu, pandaanktnya, poll, pudata, jlnuripsumoharjo, hingga, disuru, pryek, applause, bayarnya, buat, salut, stabil, suruh, indosatcare, panjaaang, sumurantri, arjuno, iso, kerjo, menghentikan, rungku, begadang, hasilnya, ganggu, internetnya, telpnya, tuhaaaannn, itu, bahkti, beraktivitas,antisipasi, dtng, habks, isi, penjual, belum, iamy, kah, dikonfirmasi, grati, kabarnya, mdahn, menabrak, palang, tdi, lainkarena, memuaskan, pasien, pindah, cont, lamasebelah, diputus, maunya, nyambung, dbantu, wonokromo, akhir, bahari, kecoklatan, gunung, besok, selasa, api, dhoho, duech, gerbong, jurusan, matisumuk, penataran, tagung, bebebrapa, bnowo, ksh, sdah, gadel, infodan, karangpoh, maintenancemohon, nyalakantrmksh, totalapakah, blas, melekan, merem, nyamuk, raiso, eksekusi, latah, ompong, perda, riil, sekedar, seok, terseok, wawasanssf, dikontak, tany, trs, bert, garasitetapi, setuju, sampahnya, beladiri, kemampuan, laksanakan, latihan, personil, polres, tingkatkan, bersama, keamanan, berlubang, jlketintang, trirismaharini, allah, bobo, kasur mabok, harapan, diblkg, kendangsari, dbantuurgent, gbs, intdk, kejelasantlg, line, kaliwaron, kog, berjam, ketahan, lokasinya, pasuran, detikdotcom, kelip, lebih, hehehe, jalanterima, gagal, mengelola, ruangkarena, sektor, tata, dirubah, halaman, konfirmdikomplain, profesional, radio, *retweeted*, wagetaman, salah, sekalian, buk, skalian, suda, amblesnya, spd, kulon, puri, yamohon, trus, ditindaklanjuti, karangpilang, surabayamohon, warugunung, toko, nggeh, nopo, sambiroto, wonten, alatbpembayaran, erroroperator, mengoperasikan, transaksi, pendukungnya, deh, jamjam, kebiasaan, amble, kauman, pdamnya, evaluasinya, ketahuan, nyolong, balas, klumprik, agt, juli, kal, banjir, maceeeet, turi, puskesmas, macetan, eduaubdedubufeduaubcedubfubd, negeri, reward, diarea, jelek, jga, telkomlsl, bening, minimal, arua, jokjamertoyudan, magelang, trimsrssbayanewsupdate, tahu, daun, dinoyo, keputran, kliatan, pran, trtutup, bener, kud, ngene, sisan, stidaknya, wez, efisien, mari, waktu, tenngilis, akhirn, jlr, pantura, perbaikanrecycling, dikerjakan, pelebarannya, sempit, diproses, indiehome, inthanks, lap, modem, ylaporan, berapa, gedangansidoarjoada, nagoya, regency, eduaubdedubueduaubdedubueduaubdedubu, terhormat, istimewa, mhon, mngarah, kwh, bpk, jualan, pnh, rangkah, bandungsurabaya, adaapa, bgini,

mlam, adha, daging, dipondok, identik, idul, matiopo, ngenteni, pembagian, qurbanpagi, sdh, ditambahcuma, harilama, loketnya, nunggunyaperlu, jerrypurbakaro, burung, cantolan, greget, kandang, kedipkediplos, bhaktisurabaya, dear, dipasang, tolongbsegera, listriks, deadline, lambat, pengaman, pengerjaannya, tnp, jaam, satu, sebab, berturut, bangkalan, diatur, odphm, geb, super, awal, berhasil, dijebol, guna, memasukkan, selang, jwb, pelaporannya, pelay, kebocoran, mgkn, umbulan, berdoa, bsa, pasrah, slama, banyaaak, gunungsarisatelit, bandaramacetalasan, dsalahkan, noted, hajatanresepsi, larangan, mengganggu, penggunaan, raperda, bulan, terselesaikan, jomblo, kakak, pemandangan, saluran, tepi, warna, warni, hilir, simo, sukomanunggal, petemon, membant, disayangkan, majunya, megah, pembenahan, ruwet, sgr, sumpeg, mendada, disitu, crew, kecoklat, keru, klau, semangat, jelas, memang, kulontolong, meluncur, wonorejomanukan, kad, beraktivita, dikrrjakan, driyorejo, lajurhrusnya, randegan, ssatu, asta, boteng, kencana, perubahan, semudah, srg, teorinya, terla, ptc, jalannyajdi, jmbatan, lgi, marvel, stlah, pemadam, disekitarnya, disusahin, gemeeeesss, kembar, mulu, nyusahiin, fly, over, ditutuputk, hak, mencla, mencle, solusics, infonyaterim, wiyungmenganti, flying, jalanmulai, margomulyo, alangkah, bagus, baiknya, dibarengi, ide, mema, menyediakan, moda, prioritas, tranportasi, menyapaapa, kotor, sembarangan, laporkan, jabon, aireduaubdedubuaaeduabdedubuaa, petug, infojalantol, sajaantrian, cowek, pombensin, keindahan, kita, luas, setujubiar, dibuatkan, reda, tunjangan, ist, keren, plastik, tissue, ajang, pengadaan, pns, dit, kombes, mduryat, hahaha, kecele, sukoasri, tidaj, bbrapa, kembalian, krban, lakukan, pembyarn, prahnya, sllu, ternyata, persed, selesaimhn, tawangsari, bensin, ngaban, pom, tanggulangin, nyalanya, belokan, dibahu, disekitar, pasurua, raci, apaterima, infonyadaerah, jlraya, listriknyaada, wiguna, jerawat, terurai, atap, panashujan, separator, jernih, keluarnya, kapolsek, kompol, rashad, berkali, hub, jawaban, pengajuan, infogempa, kepanikan, menimbulkan, kebonsari, ledakan, terdengar, umumsatu, fiber, maticheckposting, dadakan, kerto, siwalan, allhamdulillah, mintai, mrasa, pabrik, risih, tengger, berdisplin, croos, menyebrangpun, zebra, informasinya, tumpangi, gledek, diberlakukan, sempit, merambatthx, jamini, kecilpayah, padamudah, endapa, budi, karya, menteri, mudik, perhubungan, sumadi, buih, delta, kalimas, photo, sempet, damai, sekolahan, pemk, ditutu, sedatimakasih, bandar, jombang, mulyo, yakira, kedamean, hayoribet, matiin, siapayang, tuhpadahal, tangga, network, menara, sutet, tertutup, trn, kompensasinya, koordinasikan, std, pantau, sier, square, asem, kalangan, kepanasan, kertanegara, rewel, tahap, colek, cumaaaa, ngerjakan, meluber, sungainya, manual, setau, pavedment, peningkatan, rigid, sbymanyar, tipis, light, trafic, hrckckck, hrudah, nunggunyapdhl, tgg, menghindari, sebisa, mengatur, aksi, komplek, tpatn, dipekso, seng, tok, nol, tidak, memaksa, tcash, teknologi, wtcsurabaya, deltasari, penduduk, asemrowo, Impunya, trafonya,

balok, dringu, gula, wonolangan, kletek, daya, faktor, penentuan, berjarak, pegawai, pendudukwalimurid, terangan, asapnya, lalulintas, brigjend, ngingas, bambu, dibetulkan, hutan, nglewer, rendah, isu, penerbangan, serius, mtrnwn, berjenis, suv, buktinya, menunggu, profesionalitas, selam, beritaberita, gerakan, nasional, sini, bngkul, dimanjakan, dkppun, etc, kebrsihan, kwn, portable, tmn, toilet, tindaklanjut, kap, binangun, knp, makarya, tirta, anermh, gapura, mangga, sam, wtan, lis, kartini, mene, menghiraukan, nabrak, ragu, demand, grup, hbo, indo, knpa, teks, tnpa, nikah, raisa, anomali, atm, pulihkan, infoini, hemat.

Lampiran 2. Pemindaian Grid dengan teknik Stratified Cross Validation

```

Support Vector Machines with Radial Basis Function kernel

1350 samples
 110 predictor
   2 classes: 'negatif', 'positif'

No pre-processing
Resampling: Cross-validated (10 fold)
Summary of sample sizes: 1215, 1215, 1215, 1215, 1215, 1215, ..
.
Resampling results across tuning parameters:

sigma    C          Accuracy      Kappa
 0.01    0.01    0.8629629630  0.4936951134
 0.01    0.10    0.8607407407  0.4881390552
 0.01    1.00    0.8659259259  0.5039141338
 0.01   10.00    0.8881481481  0.6111306607
 0.01  100.00    0.8859259259  0.6046396068
 0.01 1000.00    0.8866666667  0.6064290674
 0.10    0.01    0.8896296296  0.5989997348
 0.10    0.10    0.8896296296  0.5989997348
 0.10    1.00    0.8896296296  0.5989997348
 0.10   10.00    0.8903703704  0.6027012739
 0.10  100.00    0.8903703704  0.6027012739
 0.10 1000.00    0.8903703704  0.6027012739
 1.00    0.01    0.8911111111  0.6032110472
 1.00    0.10    0.8911111111  0.6032110472
 1.00    1.00    0.8911111111  0.6032110472
 1.00   10.00    0.8911111111  0.6032110472
 1.00  100.00    0.8911111111  0.6032110472
 1.00 1000.00    0.8911111111  0.6032110472
10.00    0.01    0.8911111111  0.6032110472
10.00    0.10    0.8911111111  0.6032110472
10.00    1.00    0.8911111111  0.6032110472
10.00   10.00    0.8911111111  0.6032110472
10.00  100.00    0.8911111111  0.6032110472
10.00 1000.00    0.8911111111  0.6032110472
100.00    0.01    0.8911111111  0.6032110472
100.00    0.10    0.8911111111  0.6032110472
100.00    1.00    0.8911111111  0.6032110472
100.00   10.00    0.8911111111  0.6032110472
100.00  100.00    0.8911111111  0.6032110472
100.00 1000.00    0.8911111111  0.6032110472
1000.00    0.01    0.8911111111  0.6032110472
1000.00    0.10    0.8911111111  0.6032110472
1000.00    1.00    0.8911111111  0.6032110472
1000.00   10.00    0.8911111111  0.6032110472
1000.00  100.00    0.8911111111  0.6032110472
1000.00 1000.00    0.8911111111  0.6032110472

Accuracy was used to select the optimal model using the largest
value.
The final values used for the model were sigma = 1000 and c = 0
.01.

```

Lampiran 3. Keterangan Model SVM

```

Keterangan Model :
Support Vector Machine object of class "ksvm"

SV type: C-svc (classification)
parameter : cost C = 0.01

Gaussian Radial Basis kernel function.
Hyperparameter : sigma = 1000

Number of Support Vectors : 849

Objective Function Value : -5.3688
Training error : 0.2
Probability model included.

Nilai  $\alpha_i$  :
      [,1]
[1,] 0.00375
[2,] 0.00375
[3,] 0.00375
[4,] 0.00375
[5,] 0.00375
[6,] 0.00375
      :
[849,] -0.010000

Nilai Support Vector:
      [,1]
[1,] 0.00375
[2,] 0.00375
[3,] 0.00375
[4,] 0.00375
[5,] 0.00375
[6,] 0.00375
      :
[849,] -0.010000

Nilai x:
      i listrik      padam      kota      parkir      daerah      interne      moho
      t          n          box
2      0.000      0.000      3.114      2.822      0.000      0.000      0.000      0.000
3      0.000      1.533      0.000      0.000      1.846      0.000      0.000      0.000
4      0.000      0.000      0.000      0.000      0.000      2.846      1.846      0.000
7      0.000      0.000      0.000      0.000      0.000      0.000      1.846      0.000
8      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000
2
2      0.000      0.000      0.000      0.000      0.000      0.000      0.000      0.000
:

```

i	masuk	kabel	kebaka ran	suram adu	tol	padat	pdam	pipa
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	2.498	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	1.540	1.076	1.540	0.757	0.000	0.000	0.000
2								
2	0.000	0.000	0.000	0.000	0.757	0.000	0.000	0.000
:								
i	proyek	bocor	jalan	kali	cctv	sultan	jalur	mala m
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2								
2	0.000	0.000	0.814	0.000	0.000	0.000	0.000	0.000
:								
i	sms	terima	berhen ti	prata ma	total	lampu	mati	timur
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2								
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
:								
i	biar	masyara kat	pemko t	air	airnya	barat	macet	indust ri
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2								
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.463
:								

			perbaik					babat
i	satelit	arus	an	toll	dupak	lancar	ramai	an
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2								
2	1.540	0.000	0.000	0.000	0.000	0.000	0.000	0.000
:								
		peruma						
i	indah	han	jam	min	ini	jaya	etoll	gate
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2								
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
:								
		semolo				menan		
i	petugas	waru	ditutup	etol	juanda	ggal	arah	bikin
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2								
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
:								
								meng
i	exit	krian	pln	selesai	wiyung	minggu	pagi	alir
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2								
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
:								

				jaringa	pemada		jemba	
i	mobil	gerbang	dgn	n	man	taman	tan	nyala
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2								
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
:								
i	pondok	coba	kartu	kawan	gang	lalin	info	turun
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2								
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
:								
	manuka						kawas	
i	n	kasih	menit	raya	parah	perum	an	pake
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2								
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
:								
				gudan	kerahka			
i	bandara	telkom	agung	g	n	pmk	royal	unit
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2								
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
:								

	wadung			kemari		
i	asri	bantu	rumah	n	hujan	hati
2	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000
2						
2	0.000	0.000	0.000	0.000	0.000	0.000

Nilai bias:
-0.9961187

Lampiran 4. Prediksi data *testing*

<i>Confusion Matrix SVM:</i>				
	prediction			
actual	negatif	positif		
negatif	119	1		
positif	10	20		
	accuracy	precision	recall	
negatif	0.9266667	0.9224806	0.9916667	
positif	0.9266667	0.9523810	0.6666667	

<i>Confusion Matrix NN:</i>		
	prediction	
actual	negatif	positif
negatif	119	1
positif	29	1

Lampiran 5. Training Model Neural Network

10 fold stratificate cross validation

hidden: 71	thresh: 0.05	error: 717.953	time: 0.14 secs
hidden: 71	thresh: 0.05	error: 718.169	time: 0.12 secs
hidden: 71	thresh: 0.05	error: 719.702	time: 0.09 secs
hidden: 71	thresh: 0.05	error: 720.139	time: 0.11 secs
hidden: 71	thresh: 0.05	error: 722.350	time: 0.09 secs
hidden: 71	thresh: 0.05	error: 720.333	time: 0.11 secs
hidden: 71	thresh: 0.05	error: 718.113	time: 0.11 secs
hidden: 71	thresh: 0.05	error: 721.278	time: 0.09 secs
hidden: 71	thresh: 0.05	error: 720.312	time: 0.12 secs
hidden: 71	thresh: 0.05	error: 720.995	time: 0.09 secs

hasil prediksi metode Neural Network

Sampel	Negatif	Positif	Sampel	Negatif	Positif
9	0.999	0.284	705	0.999	0.914
12	0.989	0.265	722	0.994	0.917
16	0.866	0.913	731	0.997	0.431
17	0.991	0.038	751	0.995	0.996
19	0.985	0.991	797	0.989	0.423
20	0.999	0.959	806	0.993	0.766
29	1.000	0.352	819	0.993	0.766
33	0.993	0.292	825	0.996	0.759
52	0.999	0.694	830	0.806	0.987
61	0.786	0.525	861	1.000	0.998
62	0.999	0.914	866	1.000	0.909
71	0.976	0.272	886	0.991	0.919
79	1.000	0.990	891	0.999	0.997
112	0.961	0.407	893	0.990	0.932
134	0.999	0.969	895	0.999	0.893
139	0.996	0.805	900	1.000	0.988
156	0.998	0.209	909	0.991	0.740
168	1.000	0.999	917	0.919	0.380
181	1.000	0.964	920	0.993	0.049
196	1.000	0.220	943	1.000	0.954
200	0.999	0.893	959	0.970	0.993
212	0.999	0.999	969	0.996	0.007
219	1.000	0.943	974	0.987	0.137
223	0.997	0.861	991	0.993	0.766
246	0.996	0.243	995	0.989	0.988
263	0.993	0.766	998	1.000	0.985
282	0.993	0.766	1014	0.996	0.211
285	0.700	0.060	1017	1.000	0.996
293	0.993	0.766	1063	0.989	0.997
306	0.999	0.677	1075	0.993	0.867
310	1.000	0.970	1089	0.999	0.863
320	0.996	0.243	1124	1.000	0.951

347	1.000	0.438		1126	1.000	0.494
353	1.000	0.425		1127	0.993	0.766
354	0.976	0.272		1129	0.993	0.766
360	0.999	0.834		1139	0.996	0.299
362	0.990	0.255		1142	0.999	0.807
363	0.999	0.844		1153	0.999	0.858
365	1.000	0.982		1204	1.000	0.999
374	0.978	0.994		1211	0.982	0.594
416	0.993	0.766		1217	1.000	0.162
417	0.993	0.766		1229	1.000	0.951
424	1.000	0.993		1231	0.969	0.649
433	0.993	0.766		1236	0.998	0.994
439	0.996	0.122		1257	1.000	0.752
442	0.993	0.766		1261	0.994	0.147
452	0.972	0.030		1263	1.000	0.996
457	0.993	0.766		1277	0.999	0.677
479	0.993	0.962		1287	0.967	0.954
507	1.000	0.195		1295	0.999	0.760
508	0.999	0.345		1303	1.000	0.982
524	0.999	0.127		1308	0.999	0.493
528	1.000	0.713		1311	0.993	0.022
532	0.996	0.863		1316	0.993	0.115
544	0.994	0.648		1318	0.999	0.325
549	0.998	0.580		1325	0.995	0.091
551	1.000	0.996		1330	1.000	0.774
552	0.993	0.766		1333	0.993	0.766
558	1.000	0.013		1345	1.000	0.967
571	0.993	0.147		1349	1.000	0.994
578	0.979	0.257		1351	0.997	0.431
593	1.000	1.000		1367	0.995	0.971
595	0.936	0.911		1396	0.996	0.776
597	1.000	0.340		1415	0.979	0.464
602	1.000	0.992		1423	1.000	0.929
616	0.999	0.027		1430	0.999	0.992
643	1.000	0.558		1434	0.973	0.053
664	0.998	0.742		1447	1.000	0.346
673	1.000	0.690		1457	1.000	0.954
679	0.970	0.313		1460	0.999	0.877
681	0.998	0.256		1466	1.000	0.870
682	1.000	0.007		1471	0.948	0.722
685	0.999	0.990		1475	0.984	0.336
698	0.988	0.973		1482	0.993	0.766
702	0.993	0.766		1500	0.993	0.867

Lampiran 6. Bobot Neural Network untuk Neuron ke-1 di dalam hidden layer.

Variabel	Nilai		
error	720.9951	barat.to.1layhid1	-0.4258
reached.threshold	0.0000	macet.to.1layhid1	0.5988
steps	1.0000	industri.to.1layhid1	-1.6135
Intercept.to.1layhid1	-0.0747	satelit.to.1layhid1	0.2961
listrik.to.1layhid1	-1.0072	arus.to.1layhid1	-0.9243
padam.to.1layhid1	-1.0528	perbaikan.to.1layhid1	0.2031
kota.to.1layhid1	0.4791	toll.to.1layhid1	-0.7931
parkir.to.1layhid1	-0.9575	dupak.to.1layhid1	-3.4316
daerah.to.1layhid1	0.1984	lancar.to.1layhid1	0.3658
internet.to.1layhid1	-0.3342	ramai.to.1layhid1	0.5922
mohon.to.1layhid1	-0.1743	babatan.to.1layhid1	-0.2766
box.to.1layhid1	-0.1695	indah.to.1layhid1	0.6340
masuk.to.1layhid1	0.8512	perumahan.to.1layhid1	0.1328
kabel.to.1layhid1	0.5693	jam.to.1layhid1	-0.5201
kebakaran.to.1layhid1	0.5575	min.to.1layhid1	0.5449
suramadu.to.1layhid1	-0.9432	ini.to.1layhid1	-0.4239
tol.to.1layhid1	0.4620	jaya.to.1layhid1	-1.0737
padat.to.1layhid1	-1.8674	etoll.to.1layhid1	0.8317
pdam.to.1layhid1	1.6547	gate.to.1layhid1	0.6292
pipa.to.1layhid1	1.3544	petugas.to.1layhid1	0.8344
proyek.to.1layhid1	-1.3556	semolowaru.to.1layhid1	0.4944
bocor.to.1layhid1	0.3396	ditutup.to.1layhid1	0.5012
jalan.to.1layhid1	-1.4891	etol.to.1layhid1	-1.1354
kali.to.1layhid1	1.1781	juanda.to.1layhid1	0.0469
cctv.to.1layhid1	0.9868	menanggal.to.1layhid1	0.4147
sultan.to.1layhid1	1.7261	arah.to.1layhid1	-0.6498
jalur.to.1layhid1	0.3829	bikin.to.1layhid1	-0.1400
malam.to.1layhid1	1.4077	exit.to.1layhid1	1.1263
sms.to.1layhid1	0.3881	krian.to.1layhid1	0.3261
terima.to.1layhid1	0.7391	pln.to.1layhid1	-0.2710
berhenti.to.1layhid1	0.1143	selesai.to.1layhid1	-0.1649
pratama.to.1layhid1	-0.3464	wiyung.to.1layhid1	-1.7939
total.to.1layhid1	-0.4965	minggu.to.1layhid1	-0.0294
lampu.to.1layhid1	0.8497	pagi.to.1layhid1	-0.0288
mati.to.1layhid1	1.2620	mengalir.to.1layhid1	-1.3695
timur.to.1layhid1	2.5309	mobil.to.1layhid1	-1.8473
biar.to.1layhid1	-0.2754	gerbang.to.1layhid1	-0.3910
masyarakat.to.1layhid1	2.1893	dgn.to.1layhid1	-1.0736
pemkot.to.1layhid1	0.1260	jaringan.to.1layhid1	-1.1662
air.to.1layhid1	-0.0688	pemadaman.to.1layhid1	0.1862
airnya.to.1layhid1	0.3008	taman.to.1layhid1	-0.2484
		jembatan.to.1layhid1	0.6933

nyala.to.1layhid1	-0.4628
pondok.to.1layhid1	0.5678
coba.to.1layhid1	-0.7186
kartu.to.1layhid1	1.0910
kawan.to.1layhid1	-0.1537
gang.to.1layhid1	-1.9764
lalin.to.1layhid1	-0.3755
info.to.1layhid1	0.0019
turun.to.1layhid1	-1.6895
manukan.to.1layhid1	-0.3529
kasih.to.1layhid1	0.7455
menit.to.1layhid1	-0.7460
raya.to.1layhid1	-0.3819
parah.to.1layhid1	1.6110
perum.to.1layhid1	1.9112
kawasan.to.1layhid1	1.2985

pake.to.1layhid1	-0.5034
bandara.to.1layhid1	0.1692
telkom.to.1layhid1	0.3132
agung.to.1layhid1	1.5529
gudang.to.1layhid1	0.0084
kerahkan.to.1layhid1	-0.6052
pmk.to.1layhid1	0.7756
royal.to.1layhid1	0.4670
unit.to.1layhid1	-0.8321
wadungasri.to.1layhid1	0.8625
bantu.to.1layhid1	0.4162
rumah.to.1layhid1	-0.8863
kemarin.to.1layhid1	0.2724
hujan.to.1layhid1	-0.3572
hati.to.1layhid1	0.6040

Lampiran 7. Sepuluh klaster yang dibangun berdasarkan kandungan sentimen negatif

klaster1: air pdam mati daerah wiyung babatan mohon mengalir air nya barat pipa info pagi perum pondok kemarin wilayah jam perumahan pratama total ya bocor hr graha indah jalan km taman nyala jl timur box minggu no masuk bayar raya gang gg jaya lg malam d p erbaikan gangguan pemadaman ini dgn dg kawasan semolowaru petugas gk kawan lancar macet tlg tp min lampu x rumah selatan proyek cc gresik gudang kebakaran tol listrik menanggal padam pln jaringan bandara juanda jalur bs etol hujan gerbang exit sultan suramadu gate kartu ditutup arah padat parah plaza tunjungan pake parkir agung industri kerahkan mobil pmk pt royal unit wadungasri b ermasalah kabel tl jembatan jd satelit telkom

klaster2: jam internet bs parkir mohon telkom pdam pagi kebakaran min ditutup toll d mati tl no juanda etol jaringan pipa etoll manukan ini jalur air box proyek gk timur dg ya mobil jd pemkot gudang babatan arah pondok perbaikan menanggal tp hr dgn cc gang semoga telpon gangguan perumahan indah mengalir minggu suramadu padat x jembatan terima lg kali airnya bandara kemarin hujan nya la pake barat petugas kabel jl kartu bermasalah dinas bocor tlg gate masuk daerah rumah kawan bikin bayar satelit banget lancar exit total selatan kasih krian padam info wiyung gerbang sultan taman semolowaru plaza gg menit perum malam pemadaman km raya pt royal dupak bendul gresik tol macet

klaster3: listrik padam daerah jam mohon mati pemadaman nyala info kawasan pln ya semolowaru perumahan wiyung plaza tunjungan kasih x pratama jaya timur bendul perbaikan perum kabel menit manukan krian wilayah pondok jl jaringan babatan indah hujan taman barat terima telpon menanggal min selatan raya malam kebakaran km rumah cc gang dinas gangguan ini juanda air no bs jalan pagi kali mengalir total gg kawan graha d gudang tol macet tlg tp bandara box kemarin jalur hr minggu pdam etol dgn dg gerbang exit sultan suramadu gate masuk kartu ditutup arah padat parah lampu bayar pake parkir proyek agung industri kerahkan mobil pmk pt royal unit wadungasri pipa bermasalah petugas gk tl

klaster4: macet arah jembatan perbaikan parah jalan gate tl graha tol padat krian kebakaran jam timur box jalur air pdam dgn pipa jl juanda hr wiyung masuk lampu proyek bermasalah etoll kali toll perumahan bandara babatan mati total ya km daerah jaya kawan malam lancar gresik menanggal padam info indah minggu exit suramadu pagi raya petugas menit lg manukan bikin d gudang pondok listrik pln tlg tp mohon gangguan jaringan pemadaman ini min kemarin mengalir no bs etol pratama hujan dg gerbang sultan taman kartu ditutup kawasan semolowaru nyala plaza tunjungan x bayar pake rumah selatan parkir barat agung cc industri kerahkan mobil pmk pt royal unit wadungasri gang gg gk kabel

klaster5: tol gerbang masuk kartu juanda dgn gate kawan kebakaran bandara menanggal exit km taman padat petugas satelit gresik jl ya bayar macet tlg jalur bs etol parah etoll lancar arah mobil d gudang pondok perbaikan listrik padam pln tp info jam mohon gangguan jaringan pemadaman perumahan ini timur min babatan box indah kemarin mengalir air hr wiyung mati minggu no pdam pratama jalan hujan dg total sultan suramadu pagi ditutup kawasan daerah semolowaru lampu nyala plaza tunjungan x pake rumah selatan parkir proyek raya barat agung cc industri kerahkan mobil pmk pt royal unit wadungasri pipa bermasalah gang gg gk jaya kabel tl jembatan jd telkom internet menit semoga perum dinas dupak

klaster6: lampu mati mohon nyala info daerah padam menit jam tl pln kawasan ya gg semoga selatan jaya tp jaringan timur min masu k barat cc petugas perum manukan krian d jl macet tlg kemarin mi nggu no bs hujan pagi arah rumah bermasalah jd dinas banget lg m alam tol perbaikan indah jalan total parah x raya gang telkom in ternet kawan kasih terima bendul gudang kebakaran pondok listrik menanggal gangguan pemadaman perumahan bandara ini juanda babata n box mengalir jalur air hr wiyung pdam etol pratama dgn dg gerb ang exit sultan suramadu gate km taman kartu ditutup padat semol owaru plaza tunjungan bayar pake parkir proyek agung industri ke rahkan mobil pmk pt royal unit wadungasri pipa

klaster7: jalan proyek raya macet arah jl box perbaikan listrik wiyung ditutup mohon tol timur indah bs lampu kabel banget bikin pemkot airnya kebakaran masuk graha info jam pdam suramadu padat daerah parah rumah barat petugas dinas kali malam bocor gresik d tp jalur hujan parkir industri mobil satelit lancar gudang pondo k menanggal padam pln tlg gangguan jaringan pemadaman perumahan bandara ini juanda min babatan kemarin mengalir air hr mati ming gu no etol pratama dgn dg gerbang exit total sultan ya gate km t aman kartu pagi kawasan semolowaru nyala plaza tunjungan x bayar pake selatan agung cc kerahkan pmk pt royal unit wadungasri pipa bermasalah gang gg gk jaya tl jembatan jd

klaster8: pln padam d lampu daerah listrik mati jam mohon kawan tlg pemadaman menanggal hujan pagi gang jaya tp ini nyala petuga s gg gk manukan gangguan jaringan timur hr dg kawasan bayar ruma h bendul kebakaran jl macet min babatan indah mengalir air mingg u bs total suramadu proyek bermasalah kabel lg wilayah malam gud ang pondok tol perbaikan info perumahan bandara juanda box kemar in jalur wiyung no pdam etol pratama jalan dgn gerbang exit sult an ya gate km masuk taman kartu ditutup arah padat parah semolow aru plaza tunjungan x pake selatan parkir raya barat agung cc in dustri kerahkan mobil pmk pt royal unit wadungasri pipa tl jemba tan jd satelit telkom internet menit semoga perum

klaster9: tol macet arah gerbang km masuk dupak exit padat parah etoll gate pake kebakaran satelit juanda kawan gudang pondok bs etol lampu gk info kawasan bikin gresik jalur jalan ya bayar ban get lancar d menanggal jam bandara box suramadu kartu selatan pa rkir jaya jembatan menit terima jl tp hr dg total ditutup bermas alah petugas kabel tl perbaikan listrik padam pln tlg mohon gang guan jaringan pemadaman perumahan ini timur min babatan indah ke marin mengalir air wiyung mati minggu no pdam pratama hujan dgn sultan taman pagi daerah semolowaru nyala plaza tunjungan x ruma h proyek raya barat agung cc industri kerahkan mobil pmk pt roya l unit wadungasri pipa gang gg jd telkom internet

klaster10: gudang kebakaran sultan agung cc industri kerahkan mo bil pmk pt royal unit wadungasri d pondok tol jl macet perbaikan listrik menanggal padam pln tlg tp info jam mohon gangguan jarin gan pemadaman perumahan bandara ini juanda timur min babatan box indah kemarin mengalir jalur air hr wiyung mati minggu no pdam b s etol pratama jalan hujan dgn dg gerbang exit total ya suramadu gate km masuk taman kartu pagi ditutup arah kawasan padat daerah parah semolowaru lampu nyala plaza tunjungan x bayar pake rumah selatan parkir proyek raya barat pipa bermasalah petugas gang gg gk jaya kabel tl jembatan jd satelit telkom internet kawan menit semoga perum dinas dupak etoll kasih terima

Lampiran 8. Sepuluh klaster yang dibangun berdasarkan kandungan sentimen positif

klaster1: tau toll e kecepatan pake
klaster2: Pemkot Desember CCTV Tilang B
klaster3: center call hubungi info santunanjasaraharja
klaster4: jembatan tol perbaikan masyarakat etoll
klaster5: jalan perbaikan arah kemacetan lintas
klaster6: tol lancar ramai arah jalan
klaster7: jalan diperlebar Karangmenjangan kasihan kebijakan
klaster8: arah bendungan day pagi proyek
klaster9: info coba bantu ditunggu dm
klaster10: listrik terima PLN nyala alhamdulillah

Lampiran 9. Syntax for Crawling Data

```
library(RCurl)
library(twitter)
library(NLP)
consumer_key = 'l0JRdIsAmow1KhQqvWpgXXXXX'
consumer_secret =
  'dk9ymnN6VqICBEmD0zbux7e1G9ASCaAFPeBJVVQkvCgeXXXXX'
access_token = '782XXXXX-
  JMRK9NFz2u6pArSmBoMueGdCLPvIysdz2WeiXXXXX'
access_secret = 'fBbgf3Kr6aPaAURer21U56YTcI2w8mI7MkyHdkdEXXXXX'
setup_twitter_oauth(consumer_key, consumer_secret, access_token, a
  ccess_secret)
1
Surabaya_tweets = searchTwitter("e100ss", n=100000, lang="id")
Surabaya_tweets
write.csv(twListToDF(Surabaya_tweets), file="e100ss23apr-.csv")
Surabaya_tweets(class)
```

Lampiran 10. *Preprocessing Data*

```
#install package
library(NLP); library(tm); library(corpus); library(ggplot2)
library(wordcloud); library(e1071); library(stringdist)
library(caret); library(neuralnet)

# import data
pemkot <- read.csv("D:/Data TA/datapemkot.csv", header=TRUE,
stringsAsFactors = FALSE)

pemkotCorpus <- Corpus(VectorSource(pemkot[,2]))

#Preprocessing

#Menghapus URL, rt, username, punctuation, casefolding,
removeURL <- function(x) gsub("http[^[:space:]]*", "", x)
pemkot_clean <- tm_map(pemkotCorpus, removeURL)
removeRT <- function(y) gsub("RT", "", y)
pemkot_clean <- tm_map(pemkot_clean, removeRT)
removeUN <- function(z) gsub("@\\w+", "", z)
pemkot_clean <- tm_map(pemkot_clean, removeUN)
pemkot_clean <- tm_map(pemkot_clean, removePunctuation)
pemkot_clean <- tm_map(pemkot_clean, tolower)

#Menghapus stopwords
file_stop <- file("C:/Users/HP/Desktop/stopword.txt", open =
"r")
id_stopwords <- readLines(file_stop)
close(file_stop)
id_stopwords = c(id_stopwords, "amp")
pemkot_clean <- tm_map(pemkot_clean, removewords, id_stopwords)
inspect(pemkot_clean[1:5])

#Menghapus nomor, strip white space
pemkot_clean <- tm_map(pemkot_clean, removeNumbers)
pemkot_clean <- tm_map(pemkot_clean, stripwhitespace)
inspect(pemkot_clean[1:5])

#wordcloud
wordcloud(pemkot_clean, random.order=F, min.freq=50,
max.words=Inf, colors=brewer.pal(8, "Dark2"))

#Build DTM
dtm = DocumentTermMatrix(pemkot_clean)

#Mengurangi fitur yang jarang muncul dengan teknik
removeSparseTerms
adtm <- removeSparseTerms(dtm, 0.99)

#Membuat matriks TF-IDF
dtm_Matrix = weightTfIdf(adtm, normalize = TRUE)
adtm.class = cbind(dtm_Matrix, pemkot$class)
dtmtfidf = as.matrix(adtm.class)
dtmtfidftrain.df = as.data.frame(dtmtfidf)
write.csv(dtmtfidftrain.df, file="D:/Data TA/dtmsvmtune.csv")
```

Lampiran 11. *Syntax* Membentuk Model SVM

```
#Load Data
dtmsvm <- read.csv("D:/Data TA/dtmsvmtune.csv", header=TRUE,
stringsAsFactors = FALSE)

#Stratified Cross Validation
cvIndex <- createFolds(as.factor(dtmsvm$V111), k=10)

#Tuning Model SVM
for (cvi in cvIndex){
  svm.train <- dtmsvm[-cvi,2:111]
  svm.test <- dtmsvm[cvi,2:111]
  trainX <-svm.train[,1:110]

  # Setup for cross validation
  library(e1071)
  set.seed(123)
  ctrl <- trainControl(method="cv",
                        number = 10,
                        classProbs = TRUE)

  # Grid search to fine tune SVM
  grid <- expand.grid(sigma = c(.01, 0.1, 1, 10, 100, 1000),
                     C = c(.01, 0.1, 1, 10, 100, 1000)
  )

  #Train SVM
  svm.tune <- train(x=trainX,
                   y= dtmsvm[cvi,112],
                   method ="svmRadial",
                   metric="ROC",
                   tuneGrid = grid,
                   trControl=ctrl)

  svm.tune
}

#Mendapatkan Model
for (cvi in cvIndex){
  y = as.factor(dtmsvm[-cvi,112])
  svm.model = svm(y~., data=trainX, type='C-classification',
kernel='radial',scale=FALSE, gamma=1000, cost=0.01)
}

#Prediksi Model
pred = predict.train(svm.tune, newdata = svm.test)
actual = dtmsvm[cvi,112]
table(actual, pred)
```

Lampiran 12. *Syntax* Membentuk Model NN

```
dtmsvm <- read.csv("D:/Data TA/dtmsvmtune.csv", header=TRUE,
stringsAsFactors = FALSE)

dtmsvm$negatif = c(dtmsvm$V111 == 'positif' )
dtmsvm$positif = c(dtmsvm$V111 == 'negatif')
dtmsvm$V111 <- NULL

trainingclassnn <- (dtmsvm[1:1500,2:112])

library(neuralnet)
sigmoid = function(g) {
  1 / (1 + exp(-g))
}

#Stratified Cross Validation
folds <- createFolds(trainingclassnn[,111], k = 10)

#Membangun model
for (fld in folds){
  set.seed(256)
  modelnya <- neuralnet(dtmsvm[-fld,'negatif'] + dtmsvm[-
fld,'positif']~

listrik+padam+kota+parkir+daerah+internet+mohon+box+masuk+kabel
+kebakaran+suramadu+tol+padat+pdam+pipa+proyek+bocor+jalan+kali
+cctv+sultan+jalur+malam+sms+terima+berhenti+pratama+total+lamp
u+mati+timur+biar+masyarakat+pemkot+air+airnya+barat+macet+indu
stri+satelit+arus+perbaikan+tol+dupak+lancar+ramai+babatan+ind
ah+perumahan+jam+min+ini+jaya+etoll+gate+petugas+semolowaru+diti
utup+etol+juanda+menanggal+arah+bikin+exit+krian+pln+selesai+wi
yung+minggu+pagi+mengalir+mobil+gerbang+dgn+jaringan+pemadaman+
taman+jembatan+nyala+pondok+coba+kartu+kawan+gang+lalin+info+tu
run+manukan+kasih+menit+raya+parah+perum+kawasan+pake+bandara+t
elkom+agung+gudang+kerahkan+pmk+royal+unit+wadungasri+bantu+rum
ah+kemarin+hujan+hati,
  data = dtmsvm[-fld,2:111], hidden=71,
  threshold =0.05, stepmax = 500, algorithm = 'backprop',
  learningrate=0.001, err.fct = 'sse',
  linear.output = FALSE, lifesign='full', act.fct = sigmoid)

  pr.nn <- compute(modelnya,trainingclassnn[fld,1:110])
}

#Prediksi Model NN
pr.nn$net.result
results <- data.frame(actual = dtmsvm[fld,112], prediction =
ifelse(pr.nn$net.result[,1]<0.8, 'positif', 'negatif'))
table(results)
```

BIOGRAFI PENULIS



Zakya Reyhana lahir di Ujung Pandang (sekarang disebut Makassar) pada tanggal 23 Oktober 1994. Penulis merupakan anak dari pasangan Bapak Bambang Moelyono dan Ibu Ramlah Maricar. Pendidikan formal yang pernah ditempuh penulis diawali di TK Kartika Chandra Kirana VII-1 (1999-2000). Pendidikan dasar diperoleh di SD Kartika Wirabuana I (2000-2006). Selanjutnya pendidikan menengah pertama di SMP Negeri 06 Makassar (2006-2008)

dan menempuh pendidikan menengah atas di SMA Negeri 02 Makassar (2008-2011). Setelah lulus SMA penulis melanjutkan studinya di Jurusan Statistika Fakultas Sains dan Teknologi Universitas Airlangga dan memperoleh gelar Sarjana Statistika (S.Stat) pada tahun 2016. Penulis melanjutkan studi program magister di Departemen Statistika Institut Teknologi Sepuluh Nopember pada tahun 2016. Jika ada kritik dan saran yang berhubungan dengan tesis ini dapat disampaikan melalui:

email penulis: zakyareyhana@yahoo.com

akun twitter penulis: twitter.com/zakyarr