TUGAS AKHIR – SS141501

ANALISIS *PROFILING TWEETS* PADA PEMILIHAN GUBERNUR DI JAWA TAHUN 2018 MENGGUNAKAN METODE *CLUSTERED SUPPORT VECTOR MACHINES*

M ASADUR ROFIQ
NRP 062114 4000 0097

Dosen Pembimbing
Imam Safawi Ahmad, S.Si., M.Si
Dr. rer.pol. Dedy Dwi Prastyo, S.Si., M.Si

PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2018

TUGAS AKHIR – SS141501

ANALISIS *PROFILING TWEETS* PADA PEMILIHAN GUBERNUR DI JAWA TAHUN 2018 MENGGUNAKAN METODE *CLUSTERED SUPPORT VECTOR MACHINES*

M ASADUR ROFIQ
NRP 062114 4000 0097

Dosen Pembimbing
Imam Safawi Ahmad, S.Si., M.Si
Dr. rer.pol. Dedy Dwi Prastyo, S.Si., M.Si

PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2018

FINAL PROJECT – SS141501

TWEETS PROFILING ANALYSIS ON GOVERNOR ELECTION IN JAVA 2018 USING CLUSTERED SUPPORT VECTOR MACHINES METHOD

M ASADUR ROFIQ
SN 062114 4000 0097

Supervisors:
Imam Safawi Ahmad, S.Si., M.Si
Dr. rer.pol. Dedy Dwi Prastyo, S.Si., M.Si

UNDERGRADUATE PROGRAMME
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS, COMPUTING, AND DATA SCIENCE
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2018

LEMBAR PENGESAHAN

ANALISIS *PROFILING TWEETS* PADA PEMILIHAN GUBERNUR DI JAWA TAHUN 2018 MENGGUNAKAN METODE *CLUSTERED SUPPORT VECTOR MACHINES*

TUGAS AKHIR
Diajukan untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Sains
pada
Program Studi Sarjana Departemen Statistika
Fakultas Matematika, Komputasi, dan Sains Data
Institut Teknologi Sepuluh Nopember

Oleh :
M Asadur Rofiq
NRP. 062114 4000 0097

Disetujui oleh Pembimbing:
Imam Safawi Ahmad, S.Si., M.Si (.....)
NIP. 19810224 201404 1 001
Dr. rer.pol. Dedy Dwi Prastyo, S.Si., M.Si (.....)
NIP. 19831204 200812 1 002

Mengetahui,
Kepala Departemen

Dr. Suhartono
NIP. 19710929 199512 1 001

SURABAYA, JULI 2018

ANALISIS *PROFILING TWEETS* PADA PEMILIHAN GUBERNUR DI JAWA TAHUN 2018 MENGGUNAKAN METODE *CLUSTERED SUPPORT VECTOR MACHINES*

Nama Mahasiswa	**:**	**M Asadur Rofiq**
NRP	**:**	**062114 4000 0097**
Departemen	**:**	**Statistika FMKSD-ITS**
Dosen Pembimbing	**:**	**Imam Safawi Ahmad, S.Si., M.Si**
		Dr. rer.pol. Dedy Dwi Prastyo, S.Si., M.Si

Abstrak

Penggunaan sosial media dalam percakapan politik kian meningkat drastis. Twitter sebagai media komunikasi politik akan menjadi sarana efektif untuk saling bertukar ide atau gagasan terkait pasangan calon khususnya menjelang pilkada serentak 2018 di Pulau Jawa. Salah satu alat yang dapat digunakan untuk mengekstraksi opini masyarakat pada twitter adalah analisis senTimen. Hasil analisis senTimen dapat berupa klasifikasi sebuah teks berdasarkan opini yang terkandung dalam teks atau dokumen tersebut. Dalam penelitian ini metode klasifikasi yang digunakan adalah CSVM dengan fungsi kernel linear dan RBF. Metode ini adalah pengembangan dari metode SVM yang bertujuan mengurangi beban komputasi saat digunakan pada data skala besar. Hasil dari penelitian ini menunjukkan bahwa pada Provinsi Jawa Timur, Jawa Tengah dan Jawa Barat metode CSVM menggunakan kernel linear memiliki ketepakatan klasifikasi lebih baik dibandingkan dengan menggunakan kernel RBF dengan nilai akurasi dan Time proses sebesar 100%; 97,8%;98,7%; serta 7,157 detik; 1,731 detik; dan 2,329 detik.

Kata kunci: CSVM, Klasifikasi, Twitter, Pilkada Serentak.

(Halaman ini sengaja dikosongkan)

TWEETS PROFILING ANALYSIS ON GOVERNOR ELECTION IN JAVA 2018 USING CLUSTERED SUPPORT VECTOR MACHINES METHOD

Name	**: M Asadur Rofiq**
Student Number	**: 062114 4000 0097**
Department	**: Statistics FMKSD-ITS**
Supervisors	**: Imam Safawi Ahmad, S.Si., M.Si**
	Dr. rer.pol. Dedy Dwi Prastyo, S.Si., M.Si

Abstract

In recent years, the use of social media in political conversations has increased dramatically. In Indonesia, social media, especially twitter has an important role in the 2014 presidential election that won the pair of presidential candidates Joko Widodo and Jusuf Kalla. Twitter as a political communication media will be an effective means to exchange ideas or ideas related to candidate pairs especially ahead of elections in conjunction 2018 in Java. One tool that can be used to extract public opinion on twitter is the analysis of senTiments. The results of senTiment analysis can be a classification of a text based on the opinions contained in the text or document. In this research the classification method used is CSVM with linear kernel function and RBF. This method is the development of the SVM method aimed at reducing the computational load when used on large-scale data. The results of this study indicate that in East, Central, and West Java the CSVM method using linear kernel has better classification agreement than using RBF kernel with accuracy and Time 100%; 97,8%;98,7%; also 7,157 seconds; 1,731 seconds; dan 2,329 seconds.

Keyword*: *Classification, CSVM, Pilkada Simultaneously, Twitter

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Allah SWT karena berkat rahmat dan berkat-Nya penulis dapat menyelesaikan laporan Tugas Akhir dengan judul

"ANALISIS *PROFILING TWEETS* PADA PEMILIHAN GUBERNUR DI JAWA TAHUN 2018 MENGGUNAKAN METODE *CLUSTERED SUPPORT VECTOR MACHINES*".

Penyusunan dan penulisan laporan Tugas Akhir ini tidak terlepas dari bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Orang tua penulis, Ayah Komari dan Ibu Sanah Nurchasanh, yang telah memberikan dukungan sehingga penulis dapat menyelesaikan Tugas Akhir dengan baik.
2. Bapak Dr. rer.pol. Dedy Dwi Prastyo, S.Si., M.Si dan Bapak Imam Safawi Ahmad, S.Si., M.Si selaku dosen pembimbing yang telah membimbing dan memberikan arahan serta masukan kepada penulis.
3. Bapak Dr. Ir. Setiawan, M.S. dan Bapak Dr. R. Muhammad Atok, S.Si., M.Si., selaku dosen penguji yang telah memberikan masukan untuk kesempurnaan tugas akhir ini.
4. Bapak Dr. Suhartono, S.Si., M.Sc., selaku dosen wali selama masa perkuliahan yang telah banyak memberikan saran dan arahan dalam proses belajar di Departemen Statistika FMKSD ITS.
5. Semua pihak yang telah membantu dalam penulisan laporan ini, yang tidak dapat penulis sebutkan satu per satu

Penulis sangat mengharapkan kritik dan saran untuk membuat Tugas Akhir ini lebih baik. Besar harapan penulis agar Tugas Akhir ini bermanfaat bagi seluruh pihak.

Surabaya, Juli 2018

Penulis

(Halaman ini sengaja dikosongkan)

DAFTAR ISI

Halaman

DAFTAR GAMBAR

Halaman

DAFTAR TABEL

Halaman

(*Halaman ini sengaja dikosongkan*)

DAFTAR LAMPIRAN

Halaman

(Halaman ini sengaja dikosongkan)

BAB I
PENDAHULUAN

1.1 Latar Belakang

Dalam beberapa tahun terakhir, penggunaan media sosial dalam percakapan politik kian meningkat drastis. Riset menunjukkan bahwa 22% orang dewasa ikut serta dalam kampanye politik pada twitter, facebook, dan myspace dalam bulan menjelang pemilihan umum AS 2010 (Smith, 2011). Persentase ini cenderung meningkat seiring dengan bertambahnya jumlah penduduk per tahun dan perkembangan teknologi semakin pesat dengan koneksi internet yang semakin cepat dan tersebar luas.

Penggunaan media sosial pada kampanye Presiden AS Donald Trump menjadikan twitter dan media sosial lainnya sebagai bagian dari alat yang menarik untuk kampanye politik. Strategi media sosial merupakan komponen kunci dalam kemenangan Donald Trump dalam pemilihan umum presiden AS 2016. Berdasarkan google *trend analysis* minat pengguna *online* terhadap kandidat Trump tiga kali lebih tinggi dari Clinton. Trump mempunyai empat juta pengikut twitter lebih banyak dibandingkan Clinton. Begitu juga di Indonesia, twitter memiliki peran penting dalam Pemilihan Presiden 2014 yang memenangkan pasangan calon Joko Widodo dan Jusuf Kalla (Jokowi-JK). Berdasarkan hasil riset yang dilakukan oleh detik.com bersama Kedutaan Denmark tentang dialog demokrasi di twitter terkait Pemilihan Presiden 2014 terhitung dari 4 Juni hingga 9 Juni 2014 menunjukkan bahwa persentase pendukung pasangan Jokowi-JK 9% lebih tinggi daripada pasangan Prabowo-Hatta (Kusuma, 2015). Twitter telah menjadi saluran komunikasi yang sah di arena politik yang dapat merefleksikkan kondisi politik di dunia nyata (Tumasjan, 2010).

Twitter adalah layanan *microblogging* baru yang diluncurkan pada tahun 2006 dengan 330 juta jumlah pengguna aktif rata-rata setiap bulannya pada akhir tahun 2017 (Statista, 2018). Pada twitter, setiap pengguna dapat mengirim pesan pendek

hingga 140 karakter, yang dikenal dengan sebutan "*tweet*s" yang muncul pada papan pesan publik. *Timeline* publik menampilkan kicauan semua pengguna di seluruh dunia secara *real time* dari lebih dari 1 juta kicauan setiap jamnya. Secara global media sosial ini memiliki 500 juta kicauan dikirim setiap harinya dengan jumlah pengguna bulanan sebesar 332 juta (Maulana, 2016). Twitter merupakan sebuah situs *microblogging* yang sangat populer di Indonesia. Hal ini terlihat dari jumlah pengguna twitter yang mencapai 19,5 juta pengguna dari total 330 juta pengguna di dunia (Hidayat, 2014). Pada dasarnya, ide awal dibalik *microblogging* adalah memberikan informasi status personal secara terbuka pada publik. Namun, akhir-akhir ini postingan pada twitter mencakup hampir semua topik, mulai dari berita politik sampai informasi produk dalam berbagai format seperti kalimat pendek, *link website*, dan pesan langsung kepada pengguna. Khususnya dalam bulan-bulan menjelang Pemilihan Kepala Daerah (Pilkada) Serentak 2018, isu politik jelas ada di benak seluruh pengguna twitter.

Komisi Pemilihan Umum (KPU) atau Komisi Independen Pemilihan (KIP) Provinsi dan Kabupaten atau Kota telah menetapkan pasangan calon (Paslon) peserta Pilkada Serentak 2018. Pilkada Serentak 2018 akan diikuti oleh 171 daerah, terdiri dari 17 provinsi, 115 kabupaten, dan 39 kota. Tiga provinsi diantaranya berada di Pulau Jawa yaitu Jawa Timur, Jawa Tengah dan Jawa Barat. Banyak pengamat politik mengatakan tiga daerah tersebut kemungkinan besar akan menarik perhatian publik yang melampaui wilayahnya. Pertama, karena total penduduk di tiga wilayah tersebut hampir separuh seluruh penduduk Indonesia yang jumlahnya mencapai sekitar 237 Juta per 2010 (BPS, 2010). Kedua, pemenang tiga wilayah tersebut akan cukup menentukan bagi dukungan saat Pemilihan Presiden (Pilpres) 2019 (Muzani, 2018). Pada Pilpres 2014 pemenang di Jawa Barat adalah Prabowo-Hatta dengan angka yang cukup mutlak 59,78 persen suara. Dimana setahun sebelumnya, PKS sebagai partai pendukung Prabowo-Hatta, memenangi Pilkada di wilayah ini. Begitu juga dengan Jawa Tengah, pada tahun 2013 pemenang Pilkadanya

adalah PDIP dengan pasangan Ganjar Pranowo-Heru Sudjatmoko. Setahun setelahnya, Jokowi-JK menang mutlak di provinsi ini dengan mengantongi 66,65 persen suara (Muzani, 2018). Jawa Timur memang bisa dikesampingkan. Sebab ketika Pilpres lalu, Demokrat yang memenangi Pemilihan Gubernur (Pilgub) Jatim 2013 berposisi netral, tidak mendukung siapapun. Di provinsi yang berbatasan dengan Pulau Bali itu Jokowi-JK memperoleh 53,17 persen atau 11.669.313 suara. Sedangkan, Prabowo-Hatta memperoleh 46,83 persen atau 10.277.088 suara. Melihat fakta-fakta terebut maka tidak mengherankan kalau tiga wilayah tersebut akan menjadi sorotan media massa dan menjadi topik hangat untuk diperbincangkan di media sosial menjelang pelaksanaan pilkada nanti. Twitter sebagai media komunikasi politik akan menjadi sarana efektif untuk saling bertukar ide atau gagasan terkait pasangan calon. Gagasan atau opini pengguna pada twitter dapat digali lebih lanjut untuk dipelajari sehingga dapat digunakan untuk mengetahui gambaran mengenai opini masyarakat terhadap pasangan calon yang maju pada pilkada 2018. Salah satu alat yang dapat digunakan untuk mengekstraksi opini masyarakat pada twitter adalah analisis sentimen (Pozzi, 2017).

Analisis sentimen atau disebut juga *opinion mining* adalah bidang studi yang menganalisis opini, penilaian dan emosi masyarakat terhadap suatu entitas misalnya produk, pelayanan atau isu tertentu (Liu, 2012). Hasil analisis sentimen dapat berupa pengelompokkan sebuah teks atau dokumen yang bersifat positif atau negatif berdasarkan opini yang terkandung dalam teks atau dokumen tersebut. Sebelum dilakukan pengelompokkan teks perlu dilakukan *pre-processing* data yang terdiri dari *case folding, tokenizing, stemming,* dan *stop words*. Karena, teks pada twitter biasanya berisi banyak *noise* dan bagian teks lainnya yang tidak informatif seperti tag HTML, skrip dan iklan.

Metode pengelompokkan yang dapat digunakan dalam analisis sentimen diantaranya adalah *Support Vector Machines, Naive Bayes Classifier*, dan *Clustered Support Vector Machines* (CSVM). Pada penelitian ini akan menggunakan metode CSVM.

CSVM adalah salah satu metode klasifikasi yang dikembangkan dari metode *Support Vector Machines* (SVM). Kelebihan metode ini adalah lebih efisien dan memiliki akurasi yang tinggi dibandingkan dengan menggunakan metode SVM. CSVM dapat mengurangi beban komputasi saat digunakan untuk menganalisis data *non linear* dengan skala besar dengan cara membagi data kedalam beberapa klaster, kemudian mengelompokkan data dari masing-masing klaster dengan metode SVM (Gu, 2013).

Penelitian yang pernah dilakukan mengenai penggunaan metode CSVM adalah *Clustering Supoprt Vector Machines for Unlabeled Data Classification* oleh Jiang dkk (2009). Penelitian tersebut membahas hasil analisis klasifikasi empat jenis data tidak berlabel menggunakan metode CSVM dengan tiga jenis klasifikasi SVM yaitu SVM *Linear*, Polinomial dan RBF. Penelitian tersebut memberikan hasil waktu komputasi sangat efisien dengan rata-rata nila akurasi untuk keempat jenis data tersebut adalah 95,63%; 98,97%; dan 97,89% dengan menggunakan fungsi *kernel linear*, polinomial dan RBF. Sedangkan penelitian lain berkaitan dengan analisis sentimen pilkada yang telah dilakukan diantaranya oleh Ezza (2017) mengulas analisis sentimen calon gubernur DKI Jakarta menggunakan algoritma *Naive Bayes* dan SVM. Data yang digunakan bersumber dari akun twitter calon gubernur masing-masing sebanyak 1000 *tweet* yang terbagi menjadi tiga kelas yaitu mengandung emosi positif, negatif dan tidak menunjukkan emosi apapun. Hasil penelitian tersebut menujukkan tingkat akurasi sebesar 87,80% dan 85,77% dengan menggunakan metode *Naive Bayes* dan SVM. Pada penelitian ini dilakukan analisis profilling *tweet*s *netizen* menggunakan metode CSVM yang bertujuan untuk mengetahui karakteristik calon gubernur dari masing-masing provinsi dan mengelompokkan *netizen* berdasarkan data *tweet*s terhadap masing-masing calon gubernur.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah mengetahui karakteristik yang muncul dari masing -masing pasangan calon gubernur yang akan melaju pada pilkada serentak 2018 di Pulau

Jawa dan klasifikasi pendukung pasangan calon berdasarkan respon masyarakat terhadap masing-masing pasangan calon tersebut pada media sosial twitter dengan menggunakan metode CSVM.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Mengetahui karakteristik profil calon kepala daerah berdasarkan data *tweet netizen* yang berkaitan dengan calon Gubernur Jawa Timur, Jawa Tengah dan Jawa Barat.
2. Melakukan pengelompokkan *netizen* berdasarkan data *tweets* menggunakan metode CSVM untuk masing-masing provinsi.
3. Mengetahui akurasi dan waktu komputasi hasil klasifikasi menggunakan metode CSVM.

1.4 Manfaat Penelitian

Hasil yang diharapkan pada penelitian ini adalah dapat memberikan manfaat dalam bidang klasifikasi *tweet* secara umum dengan menggunakan metode CSVM. Penelitian ini diharapkan dapat membantu membantu pihak-pihak yang ingin mengetahui kata kunci yang identik terhadap tokoh publik calon Gubernur Jawa Timur, Jawa Tengah dan Jawa Barat melalui analisis *profiling tweet*s masyarakat pengguna Twitter.

1.5 Batasan Masalah

Berdasarkan perumusan masalah yang telah ditulis diatas, maka batasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya melakukan analisis terhadap *tweet* berbahasa Indonesia.
2. Penelitian ini tidak mengatasi kata dan kalimat yang cara penulisannya tidak umum (disingkat).
3. Data *tweet* menggunakan periode masa sebelum pelaksanaan Pilkada Serentak dari 27 Maret 2018 – 26 Juni 2018.

4. Data *tweet* yang digunakan bersumber dari akun *netizen* yang mengikuti akun twitter pasangan calon gubernur. Selain itu, diasumsikan tidak mendukung pasangan calon gubernur yang melaju pada Pilkada Serentak 2018 di Provinsi Jawa Timur, Jawa Tengah dan Jawa Barat.
5. Data *tweet* yang digunakan dapat berasal dari postingan akun *netizen* sendiri ataupun akun *netizen* orang lain.

BAB II
TINJAUAN PUSTAKA

2.1 *Text Mining*

Text Mining secara luas dapat didefinisikan sebagai proses pengetahuan intensif dimana pengguna berinteraksi dengan koleksi dokumen dari waktu ke waktu dengan menggunakan seperangkat alat analisis. Dengan cara yang serupa dengan data *mining*, *text mining* berusaha mengekstrak informasi yang berguna dari sumber data melalui identifikasi dan eksplorasi dari pola menarik (Feldman, 2007).

Proses *text mining* meliputi pengumpulan informasi, pengambilan informasi, teknik penambangan data termasuk analisis asosiasi dan tautan, visualisasi dan analisis prediktif. Tujuan utamanya adalah mengubah teks (data tidak terstruktur) menjadi data (format terstruktur) untuk analisis, melalui penggunaan metode pengolahan bahasa alami (Kumar, 2013). Serangkaian aktifitas harus dilakukan dalam *text mining* agar mendapat informasi yang efisien. Aktifitas itu secara umum meliputi *pre-processing* dan *feature selection*.

2.2 *Text Pre-Processing*

Tahapan *pre-processing* data memiliki peranan yang sangat penting dan krusial dalam aplikasi *text mining*. Ini adalah tahapan awal dalam aplikasi *text mining*, dimana file teks mentah akan dirubah menjadi rangkai unit bahasa yang sangat jelas (Herbich, 2010). Adapun tahapan *pre-processing* adalah sebagai berikut:

a. *Case Folding*, adalah proses mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar (biasanya huruf kecil atau *lowercase*), dengan mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf ‘a’ sampai dengan ‘z’ yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter (Weiss, 2010).
b. *Tokenizing*, merupakan tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Sehingga

sekumpulan karakter dalam suatu kalimat akan dipecah ke dalam satuan per kata.

c. *Stemming*, merupakan tahap mencari kata dasar dari tiap kata dengan menghilangkan awalan, akhiran, sisipan, dan kombinasi dari awalan dan akhiran.

d. *Filtering*, merupakan tahap mengambil kata-kata penting dari hasil *token* menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting).

2.3 *Confix-Stripping Stemmer*

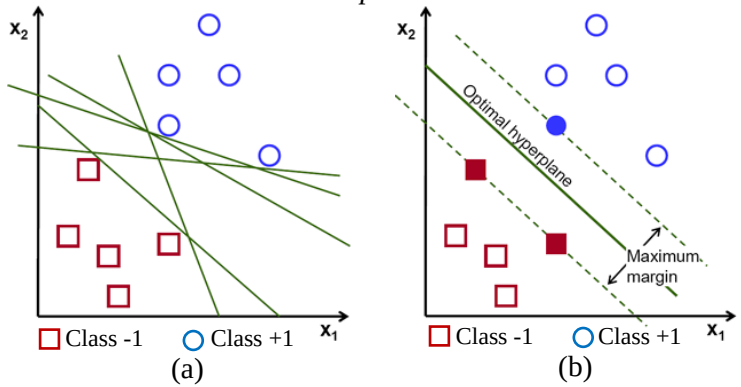
Confix stripping (CS) *stemmer* adalah metode *stemming* pada Bahasa Indonesia yang diperkenalkan oleh Jelita Asian yang merupakan pengembangan dari metode Nazief and Adriani's *Stemmer*. Proses ini berfungsi untuk mengubah bentuk dari suatu kata menjadi bentuk kata dasarnya dengan cara menghilangkan kandungan imbuhan seperti awalan dan akhiran pada kata yang bersangkutan, sehingga diharapkan diperoleh bentuk dasarnya (Arifin, 2009). Pada dasarnya, algoritma ini mengelompokkan imbuhan ke dalam beberapa kategori sebagai berikut:

1. *Inflection Suffixes* yakni kelompok-kelompok akhiran yang tidak mengubah bentuk kata dasar. Kelompok ini dapat dibagi menjadi dua:
   - *Particle* (P) atau partikel, termasuk di dalamnya adalah partikel "-lah", "-kah", "-tah", dan "-pun".
   - *Possessive Pronoun* (PP) atau kata ganti kepunyaan, termasuk di dalamnya adalah "-ku" , "-mu", dan "-nya".
2. *Derivation Suffixes* (DS) yakni kumpulan akhiran yang secara langsung dapat ditambahkan pada kata dasar. Termasuk di dalam tipe ini adalah akhiran "-i", "-kan", dan "-an".
3. *Derivation Prefixes* (DP) yakni kumpulan awalan yang dapat langsung diberikan pada kata dasar murni, atau pada kata dasar yang sudah mendapatkan penambahan sampai

dengan 2 awalan. Termasuk di dalamnya adalah awalan yang dapat bermorfologi (“me-”, “be-”, “pe-”, dan “te-”) dan awalan yang tidak bermorfologi (“di-”, “ke-” dan “se-”).

2.4 *Support Vector Machine*

Support Vector Machine merupakan algortima yang cepat dan efektif dalam masalah klasifikasi teks (Feldman, 2007). SVM pertama kali diperkenalkan oleh Vapnik pada tahun 1992 di *Annual Workshop on Computational Learning Theory* sebagai rangkaian harmonis konsep-konsep unggulan dalam bidang mengenali suatu pola. Prinsip dasar SVM adalah *linear classifier*, dan selanjutnya dikembangkan agar dapat bekerja pada *problem non-linear* dengan memasukkan konsep *kernel trick* pada ruang kerja berdimensi tinggi (Karl, Prasetyo, & Hafner, 2014). Dalam istilah geometri, pemisah biner SVM dapat dilihat sebagai *hyperlane* dalam ruang pemisah yang memisahkan ruang hal-hal positif dengan ruang hal-hal negatif. Berikut adalah gambar ilustrasi metode SVM *linear separable*.



(Sumber: https://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/ introduction to svm.html)

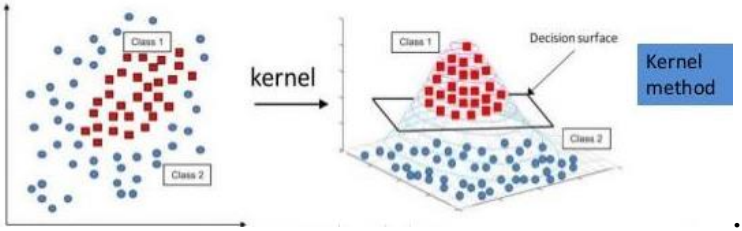
Gambar 2.1 Ilustrasi metode SVM *linear separable*

Gambar 2.1 memperlihatkan beberapa *tuples* yang merupakan anggota dari dua buah *class* yaitu: -1 dan +1. Berdasarkan tersebut dapat dilihat bahwa data 2D dapat dipisahkan secara *linier*, karena garis lurus dapat ditarik untuk memisahkan

semua anggota *class* -1 yang disimbolkan warna merah (kotak) dari *class* +1 dengan simbol warna biru (lingkaran). Terdapat sejumlah alternatif garis (*hyperlane*) pemisah tak terbatas yang dapat ditarik untuk memisahkan kedua kelas tersebut seperti ditunjukkan pada Gambar 2.1 (a). Garis solid pada Gambar 2.1 (b) menunjukkan garis pemisah terbaik yang terletak tepat pada tengah-tengah kedua class. Sedangkan *tuple fill* merah dan biru pada Gambar 2.1 merupakan *support vector*.

2.5 Fungsi *Kernel* Pada SVM

Pengelompokkan *linear* jarang sekali dijumpai dalam kasus nyata, sebab kebanyakan pada dunia *real* kasus bersifat tidak terstruktur. Untuk mengatasi hal tersebut SVM mempunyai properti yang luar biasa untuk mentransformasi menjadi *linear* dengan memasukkan fungsi *kernel*. Pada *non linear* SVM, atribut dipetakkan oleh fungsi $\Phi(\vec{x})$ ke ruang yang berdimensi lebih tinggi. SVM kemudian akan melakukan maksimum *margin* klasifikasi *linear* pada ruang tersebut terlepas bahwa data tersebut bersifat *non linear* ketika data diproyeksikan kedalam *input space*. Sehingga kedua kelas dapat dipisahkan pada ruang yang baru ini. Ilustrasi dari konsep ini dapat dilihat pada Gambar 2.2 di bawah ini.



(Sumber: https://www.slideshare.net/ankitksharma/svm-37753690)

Gambar 2.2 Fungsi Φ memetakan data ke ruang vektor

Gambar sebelah kiri (a) menujukkan bahwa data tidak dipisahkan dengan garis *linear*. Sedangkan gambar sebelah kanan (b) menunjukkan bahwa fungsi Φ memetakan setiap data pada *input space* ruang berdimensi d ($\Re^d$) ke ruang baru berdimensi

lebih tinggi $\Re^q$ setelah dilakukan transformasi *non linear*. Berikut adalah notasi matematika dari *mapping* ini.

$$\Phi : \Re^d \rightarrow \Re^q ; d < q \quad (2.1)$$

Pemetaan ini dilakukan agar dua data yang berjarak dekat pada *input space* akan berjarak dekat juga pada *feature space*, sebaliknya dua data yang berjarak jauh pada *input space* akan juga berjarak jauh pada *feature space*. Proses menemukan *titik support vector* pada SVM hanya bergantung pada *dot product* dari dua data yang sudah dilakukan transformasi pada ruang baru yang berdimensi lebih tingi yaitu $\Phi(\vec{x}_i).\Phi(\vec{x}_j)$. Sesuai teori Mercel perhitungan *dot product* dapat digantikan dengan fungsi *kernel* $K(\vec{x}_i, \vec{x}_j)$ yang mendefinisikan secara implisit transformasi Φ. Sehingga dapat dirumuskan sebagai berikut:

$$K(\vec{x}_i, \vec{x}_j) = \Phi(\vec{x}_i).\Phi(\vec{x}_j). \quad (2.2)$$

Kernel trick akan memberikan kemudahaan dalam menentukan *support vector* dalam proses pembelajaran SVM. Sebab, dengan mengetahui fungsi *kernel* untuk menentukan *support vector* tidak perlu mengetahui wujud dari fungsi *linear* Φ. Berikut ini adalah berbagai jenis dari fungsi *kernel*.

Tabel 2.1 Fungsi *Kernel* yang umum pada SVM

Jenis *Kernel*	**Fungsi**
Polynomial	$K(\boldsymbol{X}_i, \boldsymbol{X}_j) = (\boldsymbol{X}_i \cdot \boldsymbol{X}_j + 1)^h$
Gaussian Radial Basis Function (RBF)	$K(\mathbf{X}_i, \mathbf{X}_j) = e^{-\left\|\mathbf{X}_i - \mathbf{X}_j\right\|^2 / 2\gamma^2}$
Linear	$K(\mathbf{X}_i, \mathbf{X}_j) = \mathbf{X}_i{}^T \mathbf{X}_j$

2.6 *Clustering Support Vector Machines* (CSVM)

Merupakan salah satu algoritma pengembangan dari metode SVM untuk menangani data *training* dalam skala besar. Dalam pendekatan satu level CSVM, data *training* pertamakali dibagi ke dalam banyak klaster. Metode SVM diperlakukan pada masing-masing klaster untuk memodelkan hubungan *nonlinear* dalam satu

klaster. Setelah metode SVM diterapkan, sampel pengujian baru ditugaskan ke sebuah klaster berdasarkan jarak sampel klaster minimum yang didefinisikan sebagai jarak rata-rata antara sampel tersebut dan setiap sampel dalam kelompok tertentu. Semua sampel dalam penelitian ini dikodekan sebagai *vektor biner*. Jarak sampel klaster antara sampel x dan klaster tertentu adalah:

$$dist(C_i, x) = \frac{1}{n_i} \sum_{q=Ci} dist(x, q) , \tag{2.3}$$

dengan C_i adalah klaster ke i, x, sebagai sampel data, q adalah salah satu sampel dari C_i, ni adalah banyaknya sampel dari klaster C_i, dan $dist(x,q)$ adalah jarak antara sampel ke x dan q. Sebab semua fitur dari sampel x dan q adalah dalam bentuk angka kode biner, $dist(x,q)$ dirumuskan

$$dist(x, q) = \frac{Match_{11}}{Match_{11} + Match_{01} + Match_{10}} , \tag{2.4}$$

dengan *Match11* adalah jumlah fitur dimana sampel x bernilai 1 dan sampel q bernilai 1, *Match01* adalah jumlah fitur dimana sampel x bernilai 0 dan sampel q bernilai 1 dan *Match10* adalah jumlah fitur dimana sampel x bernilai 1 dan sampel q bernilai 0. Sehingga fungsi untuk menandai pengujian sampel x terhadap klaster C_j terpilih diformulasikan sebagai berikut.

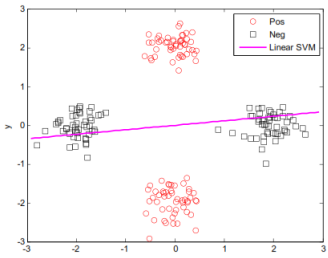
$$dist(C_j, x) = \min_{i=1,...,n} dist(C_i, x) \tag{2.5}$$

Fungsi klasifikasi SVM untuk memilih klaster C_j terhadap pengelompokan sampel x diformulasikan sebagai berikut.

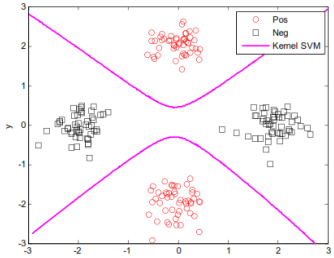
$$f_{svm\_j}(x) = (\sum_{i=1}^{sv} \alpha_i y_i K_{svm\_j}(x, x_i) + b) , \tag{2.6}$$

dengan sv adalah jumlah *support vector* dan $K_{svm\_j}(x, x_i)$ adalah fungsi *kernel* svm j *trained* untuk *cluster* C_j. Gambaran intuitif

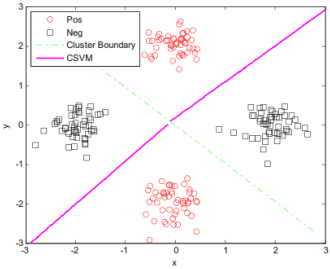
tentang *Linear* SVM, *Kernel* SVM dan CSVM bekerja untuk data *linear nonseparable*.



(a) *Linear* SVM



(b) *Kernel* SVM



(c) CSVM

(Sumber: https://www.semanticscholar.org/paper/Clustered-Support-Vector-Machines-Gu-Han/1969c621cc4d7464fc62424304ff4fb35ec72b1a)

Gambar 2.3 Fungsi memetakan data ke ruang vector

2.7 *Stratified Repeated Holdout*

Stratified hold out merupakan metode statistik yang dapat digunakan untuk mengestimasi kinerja suatu model "*machine learning*" dimana data dipisahkan menjadi dua subset yaitu data *training* dan data *testing* (Witten, Eibe, & Hall, 2011). Dalam metode *holdout*, data awal yang diberi label dipartisi ke dalam dua himpunan secara *random* yang dinamakan data *training* dan data *testing*. Proporsi data yang dicadangkan untuk data *training* dan data *testing* tergantung pada analisis misalnya 50%-50% atau 2/3 untuk *training* dan 1/3 untuk *testing*. Pada penelitian ini akan dilakukan *stratified hold out* dengan perulangan sebanyak 10 kali pengujian akurasi model. *Stratified* adalah proses pengambilan sampel secara *random* agar setiap data *training* dan *testing*

memiliki proporsi kelas yang sama dengan data *imbalanced* dan perwakilan data setiap kelas baik pada data *training* maupun data *testing*. Berikut ini adalah ilustrasi dari *Stratified Repeated Holdout dengan 6 kali perulangan*.

Gambar 2.4 Ilustrasi Partisi Data

2.8 Pengujian Performa Klasifkasi

Model klasifikasi yang telah diperoleh, selanjutnya akan diukur performa ketepatannya dalam melakukan klasifikasi. *Confusion matrix* merupakan salah satu teknik yang berguna untuk mengukur performa dari sebuah algoritma klasifikasi (Han, Kamber, & Pei, 2012). Berikut ini adalah tabel *Confusion matrix* untuk dua kelas klasifikasi.

Tabel 2.2 *Confusion Matrix*

Kelas Aktual	Kelas Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

TP adalah *True Positive*, FP adalah *False Positive*, TN adalah *True Negative*, dan FN adalah *False Negative*. Terdapat tiga jenis pengukuran yang sering digunakan dalam menghitung ketepatan klasifikasi yaitu *akurasi*, *sensitivity*, dan *specificity*.

Akurasi adalah proporsi jumlah total prediksi yang benar. Akurasi digunakan untuk menghitung ketepatan klasifikasi sebuah dokumen yang mempunyai data yang *balanced* pada tiap kategorinya. *Sensitivity* proporsi kasus positif yang diidentifikasi dengan benar. Sedangkan *specificity* adalah proporsi kasus negatif yang diidentifikasi dengan benar. Berikut adalah rumus dalam menghitung akurasi, *specivicity* dan *sensitivity*.

$$\text{Akurasi} = \frac{TN + TP}{TN + TP + FN + FP}, \tag{2.7}$$

$$Sensitivity = \frac{TP}{TP + FN}, \tag{2.8}$$

$$Specificity = \frac{TN}{TN + FP}. \tag{2.9}$$

Pada kasus data *inbalance*, pengukuran ketepatan klasifikasi dapat menggunakan nilai *G-Mean*. Nilai *G-Mean* dapat diperoleh dari hasil rata-rata geometrik nilai *recall* (Sun, Karnel, & wang, 2006). Selain itu untuk meringkas kinerja sebuah *classifier* dalam satu nilai digunakan nilai *Area Under Curve* (AUC). Berikut adalah rumus untuk mendapatkan nilai *G-Mean* dan AUC.

$$G - Mean = \sqrt{sensitivity \times specificity} \tag{2.10}$$

$$AUC = \frac{1}{2}(sensitivity + specificity) \tag{2.11}$$

2.9 Twitter

Twitter merupakan suatu layanan *social network* yang berkategori *microblogging*, dimana pengguna bisa saling bekomunikasi dalam pesan singkat berbasis teks hingga 140 karakter yang dikenal dengan *tweet*s kepada pengguna lainnya. Selain itu twitter juga dapat dikenal sebagai jejaring informasi untuk menyebarkan informasi atau berita terkini keluruh belahan dunia. Twitter menawarkan dua alat untuk membuat saling terkoneksi satu sama lain antar penggguna: *mention*(@) dan *hastags*(#). *Mention*(@) memungkinkan pengguna untuk menandai pengguna tertentu dalam *tweet*s. *Hashtag* digunakan untuk memulai, dan berpartisipasi dalam *platform* kelompok

percakapan. Dengan mengklik *hashtag* di *tweet*, pengguna bisa melihat *tweet* lain yang menggunakannya. Ini membantu dalam melacak percakapan. Sehingga dengan penggunaan *hastags* bisa diketahui topik pembicaraan yang sedang *trending* dibicarakan oleh para pengguna Twitter.

2.10 Pilkada Pulau Jawa 2018

Pilkada serentak merupakan agenda politik nasional dengan penyelenggaraan pemilihan kepala daerah yang meliputi tingkat provinsi, kabupaten dan kota dalam lingkup wilayah atau kawasan tertentu yang dilakukan secara serentak dengan tujuan terciptanya efektivitas dan efisiensi dalam pelaksanaan sehingga dapat menghemat anggaran. Komisi Pemilihan Umum (KPU) Republik Indonesia (RI) sudah menetapkan tanggal pemungutan suara Pilkada Serentak 2018 akan dilaksanakan pada tanggal 27 Juni 2018. Rencananya ada 171 daerah yang mengikuti Pilkada 2018. Tahapan Pilkada serentak 2018 akan dimulai 10 bulan sebelum hari pencoblosan. Hal ini berarti tahapan dimulai Agustus 2017. Pilkada serentak tahun 2018 akan digelar lebih besar daripada Pilkada sebelumnya. Sebanyak 171 daerah akan berpartisipasi pada ajang pemilihan kepala daerah tahun depan. Dari 171 daerah tersebut, ada 17 provinsi, 39 kota, dan 115 kabupaten yang akan menyelenggarakan Pilkada di tahun 2018. Beberapa provinsi di antaranya adalah Jawa Barat, Jawa Tengah, dan Jawa Timur. Komisi Pemilihan Umum Provinsi (KPU) Jawa Tengah telah menetapkan Pemilihan Gubernur Provinsi Jawa Tengah 2018 akan diikuti oleh dua pasangan calon, yakni Sudirman Said-Ida Fauziyah dan Ganjar Pranowo-Taj Yasin. Begitu juga Pilkada Provinsi Jawa Timur juga akan diikuti dua pasangan calon, yakni Khofifah Indar Parawansa-Emil Dardak dan Saifullah Yusuf (Gus Ipul)-Puti Guntur Soekarnoputri. Sedangkan Pilkada Provinsi Jawa Barat jumlah pasangan calon yang resmi mengikuti Pilkada Jabar 2018 adalah sebanyak empat pasangan calon, yakni Deddy Mizwar-Dedi Mulyadi, Sudrajat-Ahmad Syaikhu, Ridwan Kamil-Uu Ruzhanul Ulum, dan TB Hasanudin-Anton Charliyan.

BAB III
METODOLOGI PENELITIAN

3.1 Sumber Data

Sumber data yang akan digunakan dalam penelitian ini adalah *tweet* dari pengguna Twitter di Indonesia dengan *keywords* dari masing-masing calon yang terkumpul dari hasil *crawling* 16 akun twitter calon kepala daerah dan calon wakil kepala daerah di Provinsi Jawa Timur, Jawa Tengah dan Jawa Barat. Data tersebut diambil pada periode masa kampanye hingga menjelang pelaksanan pilkada yaitu 27 Maret 2018 – 26 Juni 2018.

3.2 Struktur Data

Struktur data yang diambil dari *website www.twitter.com* dengan bantuan Twiter API software R 3.4.3 dibuat seperti pada Tabel 3.1.

3.3 **Tabel 3.1** Struktur Data Sebelum *Preprocessing*

No.	**Nama Akun Gubernur**	***Tweet (y)***	**Nama Akun Wakil Gubernur**	***Tweet (y)***
1		y_1		y_1
2		y_2		y_2
.	@gusipul4	.	@PutiSoekarno	
.		.		
.		.		
n_1		y_{n1}		yn_2
1		y_1		y_1
2		y_2		y_2
.	@KhofifahIP	.	@EmilDardak	
.		.		
.		.		
n_3		yn_3		yn_4

No.	Nama Akun Gubernur	*Tweet (y)*	Nama Akun Wakil Gubernur	*Tweet (y)*
1		y_1		y_1
2		y_2		y_2
.	@ganjarpranowo	.	@taj_yasinmz	
.		.		
.		.		
n_5		yn_5		yn_6
1		y_1		y_1
2		y_2		y_2
.	@sudirmansaid	.	@idafauziyah	
.		.		
.		.		
n_7		yn_7		yn_8
1		y_1		y_1
2		y_2		y_2
.	@ridwankamil	.	@uuruzhan	
.		.		
.		.		
n_9		yn_9		yn_{10}
1		y_1		y_1
2		y_2		y_2
.	@MayjenSudrajat	.	@syaikhu_ahmad	
.		.		
.		.		
n_{11}		yn_{11}		yn_{12}
1		y_1		y_1
2		y_2		y_2
.	@Deddy_Mizwar_	.	@DediMulyadi71	
.		.		
.		.		
n_{13}		yn_{13}		yn_{14}
1		y_1		y_1
2		y_2		y_2
.	@tbhasanuddin	.	@AntonCharlian	
.		.		
.		.		
n_{15}		yn_{15}		yn_{16}

Struktur data pada Tabel 3.1 menunjukkan jumlah data *tweet netizen* dari masing-masing akun calon gubenur. Data *tweet netizen* tersebut kemudian diambil kata yang sering muncul dari masing-masing provinsi sebagai sebagai *profile* calon gubenur dari setiap provinsi. Pada penelitian ini variabel prediktor yang digunakan yaitu kata dasar setiap *tweet netizen* dan variabel respon yaitu klasifikasi preferensi dukungan *netizen* terhadap salah satu pasangan calon sesuai nomor urut resmi yang ditetapkan oleh KPU di setiap provinsi. Berikut ini adalah struktur data frekuensi *tweet netizen* terhadap *profile* masing-masing calon setiap provinsi.

Tabel 3.2 Struktur Data Frekuensi dan Klasifikasi *Tweet Netizen*

Netizen	Klasifikasi Pendukung	Cagub 1	Wagub 1	Cagub 2	Wagub 2	Kata 1	Kata 2	...	Kata p
1	isi_1	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	...	f_{1p}
2	isi_2	f_{21}	f_{22}	f_{23}	f_{24}	f_{25}	f_{26}	...	f_{2p}
3		f_{31}	f_{32}	f_{33}	f_{34}	f_{35}	f_{36}	...	f_{3p}
.		.	.	.	.	.	.		.
.		.	.	.	.	.	.		.
.		.	.	.	.	.	.		.
n	isi_n	f_{n1}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	....	f_{np}

Struktur data pada Tabel 3.2 menunjukkan frekuensi kata muncul pada *tweet* dari masing masing *netizen* terhadap calon gubernur untuk provinsi Jawa Timur dan Jawa Tengah. Sedangkan provinsi Jawa Barat terdapat penambahan sebanyak 4 kolom. Karena pada provinsi tersebut hanya terdapat empat pasangan calon yang melaju pada Pilkada Serentak 2018.

3.4 Langkah Analisis

Pada penelitian ini langkah analisisnya adalah sebagai berikut:

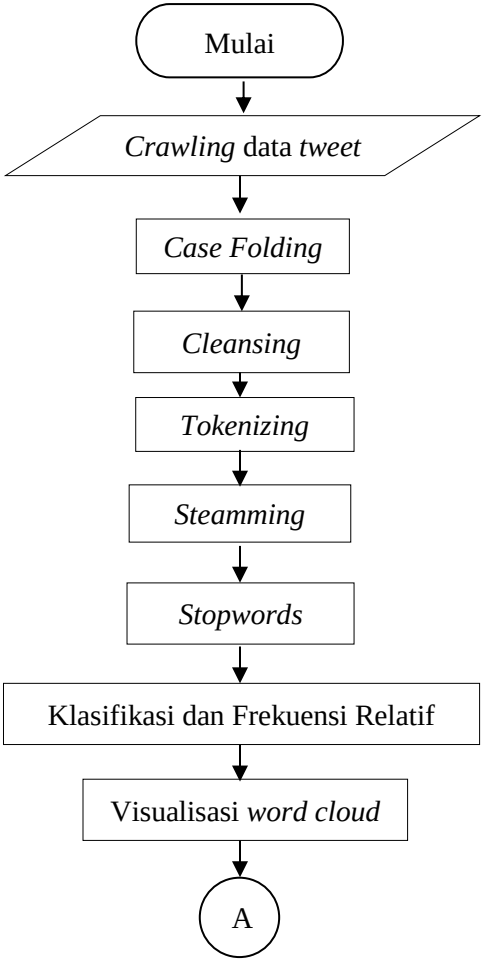
1. Mengambil data *tweet* dengan *standard search* API.
   a) Memasukkan kata kunci yaitu akun twitter pasangan calon kepala daerah yang melaju pada Pilkada Serentak 2018 di Provinsi Jawa Timur, Jawa Tengah dan Jawa Barat.

b) Menyimpan hasil *searching* ke database.

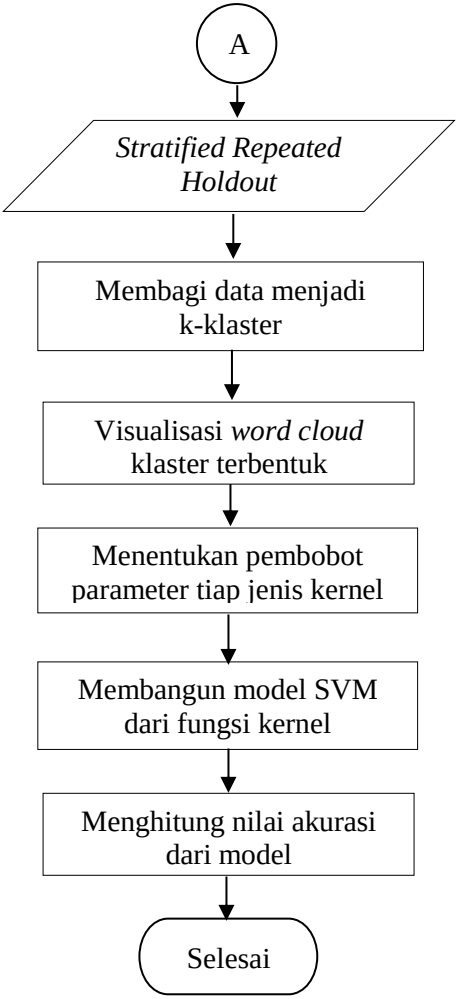
2. Menyiapkan data *tweet*, daftar *stopwords*, dan kata dasar.
3. Praproses Teks.
   a) Mengubah semua *tweet* menjadi huruf kecil. Proses ini mengkonversi data *tweet* yang berupa teks ke dalam bentuk huruf kecil. Tujuannya agar data *tweet* berada dalam bentuk konsisten.
   b) Menghapus *Emoticons* dan Tanda Caca. *Emoticons* dan tanda caca perlu dihilangkan karena tidak diperlukan dalam analisis. Data *emoticons* berubah menjadi kotak persegi ketika dilakukan ekstraksi dari twitter. Data ini akan menjadi sampah dalam dokumen jika tidak dihilangkan.
   c) Menghapus *link* URL's. Langkah ini perlu dilakukan karena *link* URL's tidak memberikan informasi apapun selama proses analisis.
   d) Menghapus simbol re*tweet* (response *tweet*) "RT" dan *hashtag*(#).
   e) Melakukan *tokenizing* untuk memecah data *tweet*s ke dalam satuan kata.
   f) Melakukan *steeming* untuk memperoleh kata dasar dengan cara menghilangkan kata imbuhan.
   g) Melakukan *filtering*. Proses ini bertujuan menghapus kata pada *tweet netizen* yang terdapat dalam daftar *stopwords*.
   h) Menentukan klasifikasi pendukung berdasarkan frekuensi kemunculan kata data *tweet*s *netizen* terhadap akun twitter pasangan calon kepala daerah.
   i) Mendapatkan frekuensi relatif dari setiap unik kata pada data *tweets* dengan kriteria frekuensi kemunculan lebih dari 20.

4. Melakukan visualisasi data *tweet* dari masing-masing pasangan calon dan provinsi dengan *word cloud*.
5. Klasifikasi teks menggunakan CSVM.
   a) Melakukan partisi data menjadi data *training* dan data *testing* menggunakan *Stratified Ten Repeated Holdout* dengan perbandingan *training-testing* sebesar 80%:20%.
   b) Membagi data *tweet* menjadi k-klaster berdasarkan jumlah pasangan calon dari setiap provinsi.
   c) Melakukan visualisasi data *tweet* dari klaster terbentuk pada provinsi Jawa Timur, Jawa Tengah dan Jawa Barat dengan *word cloud*.
   d) Menentukan pembobot parameter pada SVM tiap jenis *kernel*.
   e) Membangun model SVM menggunakan fungsi *kernel Linear* dan RBF.
   f) Menghitung nilai akurasi dari model yang terbentuk.

Diagram alir dari langkah analisis data pada penelitian ini disajikan dalam Gambar 3.1.



Gambar 3.1 Diagram Alir Praproses Teks



Gambar 3.2 Diagram Membangun Model CSVM

(*Halaman ini sengaja dikosongkan*)

BAB IV
ANALISIS DAN PEMBAHASAN

Analisis yang akan dilakukan dalam bab ini meliputi karakteristik calon kepala daerah, hasil pengelompokkan *netizen* dan klasifikasi *tweet netizen* menggunakan metode CSVM. Sebelum melakukan analisis, dilakukan terlebih dahulu *pre-processing text.*

4.1 Praproses Teks

Data *tweet netizen* mengenai calon kepala daerah di Provinsi Jawa Timur, Jawa Tengah, dan Jawa Barat yang telah terkumpul kemudian dilakukan praproses teks yaitu meliputi *case folding, tokenizing, cleansing, setemming,* dan *stopwords.* Proses tersebut dilakukan menggunakan *software jupyter notebook* dengan bahasa pemograman pyhton 2. Berikut ini merupakan struktur data *tweet netizen* mengenai calon kepala daerah Provinsi Jawa Barat yaitu Dedy Mizwar sebelum dilakukan praproses data.

Tabel 4.1 Struktur Data Dedy Mizwar Sebelum Praproses

Netizen	*Tweet*
PanjiGarutFM	Slmt pagi semangat pagi di Jum'at berkah sblm beraktifitas mari kita ngopi dulu #Patpatgulipat #DM4Jabar… https://t.co/Z7BxSOGOpT
Putri_Tuber	RT @Imambae67781157: @bekerjamelayani @Deddy_Mizwar_ @zarazettirazr @panca66 @Umar_Hasibuan_ @topelucky @DPP_Golkar
Putri_Tuber	RT @TbYusup: @bekerjamelayani @Deddy_Mizwar_ @zarazettirazr @panca66 @Umar_Hasibuan_ @topelucky @DPP_Golkar
.	.
.	.
.	.
MustikaTevia	RT @EkaPangulimaraH: Terima kasih Pak Wagub @Deddy_Mizwar_ Jumat kemarin temans bidan PTT (daerah) Prov Jabar sudah terima gaji hingga Febr…

Tabel 4.1 menunjukkan data akun dan *tweet netizen* yang belum dilakukan praproses. Data *tweet* tersebut masih mengandung simbol *retweet* (RT), tanda baca, memuat *username, link* URL, dan kata-kata yang dianggap bukan kata penting yang

berhubungan Pilkada Serentak 2018 sehingga perlu dilakukan praproses guna mendapatkan data *tweet* yang tidak memuat hal-hal tersebut. Selain itu, praproses data bertujuan untuk mengetahui karakteristik setiap pasangan calon kepala daerah dalam interval periode waktu tertentu. Berikut merupakan hasil praproses data *tweet netizen* terhadap akun twitter Dedy Mizwar.

Tabel 4.2 Struktur Data Dedy Mizwar Setelah Praproses

Netizen	*Tweet*
PanjiGarutFM	slmt pagi semangat pagi berkah sblm beraktifitas mari ngopi patpatgulipat dm4jabar
MustikaTevia	terima kasih wagub jumat kemarin temans bidan daerah prov jabar terima gaji febr
MustikaTevia	gub jabar wagub bidan ptt prov jabar terima gaji mohon
mapunk121	jabar beda jatim jateng beda
.	.
.	.
.	.
RifaldiRiedwan	bang temu tani padi tani gincu tani garam kyai majelis

Kata pada data *tweet netizen* yang telah dilakukan praproses, selanjutnya akan digunakan sebagai karakteristik calon kepala daerah dari masing-masing wilayah untuk menghitung frekuensi kata yang muncul dari *tweet* setiap akun *netizen* terhadap kata karakteristik calon kepala daerah dari masing-masing wilayah dan label preferensi dukungan *netizen* terhadap salah satu pasangan calon pada Provinsi Jawa Timur, Jawa Tengah, dan Jawa Barat. Label preferensi dukungan sesuai dengan nomor urut resmi pasangan calon kepala daerah yang telah ditetapkan oleh KPU pada setiap provinsi. Berikut adalah tabel struktur data frekuensi *tweet netizen* terhadap pasangan calon kepala daerah Provinsi Jawa Timur.

Tabel 4.3 Frekuensi Data *Tweet Netizen* Terhadap Calon Kepala Daerah Provinsi Jawa Timur

Netizen	Label	KhofifahIP	EmilDardak	Gusipul4	PutiSoekarno	Kata 1	Kata 2	...	Kata 1728
2019GNPresiden	1	1	0	0	0	0	0	...	0
22what22	2	0	0	0	2	0	0	...	0
24f81aec51b9419	1	1	0	0	0	0	0	...	0
2ayn_M	1	2	0	0	0	0	0	...	0
30jar	1	1	0	0	0	0	0	...	0
3179_k	1	1	0	0	0	0	0	...	0
3363yasri	1	2	0	0	0	1	1	...	0
350Indonesia	1	1	1	1	0	0	0	...	0
3Wulandarie	1	10	8	0	0	6	8	...	0
3anfibi	1	2	0	0	0	0	0	...	0
3mtr1	1	4	1	0	0	0	0	...	0
445a45	1	1	0	0	0	0	0	...	0
489Agus	1	1	2	2	0	0	0	...	0
4bim4nyu	1	1	0	0	0	0	0	...	0
.		.	.	.	.	.	.	.	.
.		.	.	.	.	.	.	.	.
.		.	.	.	.	.	.	.	.
zy_zzzz	1	1	0	0	0	0	0	.	0
7042		21883	14078	4657	1217	11199	8082	...	11

Kolom label pada Tabel 4.3 menunjukkan preferensi dukungan *netizen* terhadap salah satu pasangan calon berdasarkan persepsi peneliti dari frekuensi *netizen* menyebut kedua pasangan calon tersebut. Angka 1 menunjukkan bahwa frekuensi *netizen* menyebut nama pasangan calon Khofifa-Emil lebih tinggi daripada pasangan calon Gus Ipul-Puti, begitupun sebaliknya. Sepanjang periode kampanye hingga periode pelaksanaan pemungutan suara 27 Juni 2018, total terdapat 41.835 *tweet* yang berhubungan dengan kedua pasangan calon tersebut dengan jumlah *netizen* yang aktif dalam percakapan mengenai kedua pasangan calon tersebut sebanyak 7.042 *netizen*. Terdapat 21.883 *tweet* yang menyebut akun twitter pasangan calon khofifaIP dan 14078 menyebut akun twitter akun pasangan calon EmilDardak. Sedangkan kata GusIpul yang dalam hal ini sebagai akun resmi twitter calon Saifullah Yusuf muncul sebanyak 4657 kali dalam *tweet netizen* dan wakilnya PutiSoekarno disebut sebanyak 1217 kali. Selain itu dengan melihat informasi pada Tabel 4.3 dapat dilihat perbedaan aktivitas kedua pasangan calon dalam menggunakan media sosial twitter sebagai salah satu media kampanye politik. Pasangan calon Khofifah-Emil lebih sering dalam menggunakan media sosial twitter dari pada pasangan Gus Ipul-Puti sebagai salah satu sarana kampanye politik. Hal ini dibuktikan dengan jumlah akun yang menyebut pasangan Khofifah-Emil selisihnya cukup jauh terhadap pasangan Gus Ipul-Puti yaitu sebesar 30087 *tweet*. Hal ini dikarenakan tingkat popularitas pasangan Khofifah-Emil jauh lebih lebih tinggi dibandingkan pasangan Gus Ipul-Puti dalam media sosial twitter dengan dibuktikkan jumlah total *follower* pasangan Khofifah-Emil mencapai 291.910, sedangkan pasangan Gus Ipul-Putih hanya mencapai 31.859 *follower* per 29 Juni 2018. Secara keseluruhan, dari jumlah *tweet netizen* terhadap kedua pasang calon, total terdapat 1728 jenis unik kata. Frekuensi kemunculan kata paling tinggi adalah kata akun twitter Khofifah.

Selanjutnya adalah tabel frekuensi relatif data *tweet* terhadap calon kepala daerah Provinsi Jawa Timur dengan jumlah pasangan calon sebanyak 2.

Tabel 4.4 Frekuensi Relatif Data *Tweet Netizen* Terhadap Calon Kepala Daerah Provinsi Jawa Timur

Netizen	Label	KhofifahIP	EmilDardak	Gusipul4	PutiSoekarno	Kata 1	Kata 2	...	Kata 1728
2019GNPresiden	1	0.000024	0	0	0	0	0	...	0
22what22	2	0	0	0	0.000048	0	0	...	0
24f81aec51b9419	1	0.000024	0	0	0	0	0	...	0
2ayn_M	1	0.000048	0	0	0	0	0	...	0
30jar	1	0.000024	0	0	0	0	0	...	0
3179_k	1	0.000024	0	0	0	0	0	...	0
3363yasri	1	0.000048	0	0	0	0.000089	0.000124	...	0
350Indonesia	1	0.000024	0.000024	0.000024	0	0	0	...	0
3Wulandarie	1	0.000239	0.000191	0	0	0.000536	0.00099	...	0
3anfibi	1	0.000048	0	0	0	0	0	...	0
3mtr1	1	0.000096	0.000024	0	0	0	0	...	0
445a45	1	0.000024	0	0	0	0	0	...	0
489Agus	1	0.000024	0.000048	0.000048	0	0	0	...	0
4bim4nyu	1	0.000024	0	0	0	0	0	...	0
.		.	.	.	.	.	.	.	.
.		.	.	.	.	.	.	.	.
.		.	.	.	.	.	.	.	.
zy_zzzz	1	2.39E+06	0,0	0,0	0,0	0,0	0,0	.	.
7042		1	1	1	1	1	1	...	1

Setelah dilakukan perhitungan frekuensi kemunculan kata dari setiap unik kata terhadap keseluruhan *tweet* dari setiap *netizen* selama periode kampanye hingga sebelum pelaksanaan pemungutan suara terhitung sejak tanggal 27 Maret hingga 26 Juni 2018, kemudian selanjutnya dilakukan perhitungan frekuensi relatif dari setiap *netizen*. Tabel 4.4 merupakan hasil perhitungan frekuensi relatif data *tweet netizen* terhadap kedua pasangan calon kepala daerah. Perhitungan nilai frekuensi relatif setiap kata didapat dari hasil pembagian frekuensi unik kata yang muncul pada setiap *netizen* dengan frekuensi unik kata muncul dalam *tweet* semua *netizen*. Misalnya pada Tabel 4.3 *netizen* dengan akun twitter "3Wulandarie" menyebut kata akun pasangan calon Khofifah yaitu "KhofifahIP" sebanyak 10 kali dari keseluruhan *tweetnya* selama periode kampanye hingga menjelang dilakukan pemungutan suara yaitu pada tanggal 27 Maret hingga 26 Juni 2018. Sedangkan kata akun twitter "KhofifahIP" selama periode tersebut muncul dalam *tweet netizen* berulangkali sebanyak 21.883. Maka selanjutnya untuk menghitung nilai frekuensi relatif *netizen* pada akun twitter "3Wulandarie" terhadap kata "KhofifahIP" adalah dengan cara membagi nilai frekuensi kemunculan kata "KhofifahIP" pada *tweet netizen* "3Wulandarie" dengan frekuensi kemunculan kata "KhofifahIP" dari keseluruhan *tweet netizen* yaitu dalam contoh ini 10 dibagi dengan 21.883. Sehingga diperoleh nilai frekuensi relatif pada tabel 4.4 yaitu sebesar 0,000239. Begitu juga untuk untuk mendapatkan nilai frekuensi relatif pada jenis unik kata yang lainnya dari setiap *netizen*. Semakin sering unik kata muncul pada *tweet netizen*, maka nilai frekuensi relatif juga semakin besar. Hasil penjumlahan nilai frekuensi relatif dari setiap unik adalah sama dengan satu. Nilai frekuensi relatif tersebut selanjutnya akan digunakan dalam analisa klasifikasi menggunakan metode *Clustered Support Vector Machine (*CSVM).

Selanjutnya adalah tabel frekuensi data *tweet* terhadap calon kepala daerah Provinsi Jawa Tengah dengan dua pasangan calon.

Tabel 4.5 Frekuensi Data *Tweet Netizen* Terhadap Calon Kepala Daerah Provinsi Jawa Tengah

Netizen	Label	Ganjar	Taj Yasin	Sudirman	Ida Fauziyah	Kata 1	Kata 2	...	Kata 2954
1MenujuDamai	2	83	0	235	85	80	26	...	0
4Y4NKZ	1	228	2	16	1	27	9	...	0
AEPishere	1	216	0	25	5	28	3	...	0
ARIFINjogor	1	76	32	2	1	8	3	...	0
AWPramestuti	1	166	1	0	0	10	5	...	0
Addarul1	2	61	0	133	52	41	18	...	0
Adra94193051	2	41	0	136	92	66	20	...	0
AgungWidrajat	1	157	0	61	7	28	5	...	0
AhmadTa75651499	1	1	19	0	0	6	1	...	0
Alim75052441	2	14	2	135	64	39	31	...	0
AmharMaya	2	4	0	162	68	74	24	...	0
AnonymWae	2	26	0	216	161	71	56	...	0
AquinaKinanti	1	401	8	51	5	31	17	...	1
Aryprasetyo85	1	111	0	18	1	26	1	...	0
.		.	.	.	.	.	.	.	.
.		.	.	.	.	.	.	.	.
.		.	.	.	.	.	.	.	.
ztw868992	1	177	1	9	0	16	5	.	0
234		36295	439	26245	14179	10249	3965	....	20

Pada Provinsi Jawa Tengah dari total 4.630 *netizen* yang aktif terlibat dalam percakapan politik terhadap kedua pasangan calon, hanya terdapat 234 *netizen* saja yang digunakan data *tweet*nya dalam penelitian. Hal ini dilakukan untuk mengurangi beban komputasi saat dilakukan analisis. Berdasarkan Tabel 4.5 dapat diketahui bahwa pasangan calon Ganjar Pranowo-Taj Yasin lebih sering muncul dalam *tweet netizen* daripada lawan politiknya Sudirman Said-Ida Fauziyah. Sebanyak 3.6295 *tweet* yang menyebut kata akun twitter pasangan calon Ganjar Pranowo Sedangkan wakil pasangan calon Ganjar Pranowo yaitu Taj Yasin hanya muncul sebanyak 439 kali dalam *tweet netizen*, sangat jauh jika dibandingkan dengan calon lainnya. Hal ini dikarenakan jumlah *follower* yang dimiliki hanyak sejumlah 707 *netizen*. Sedangkan Ganjar Pranowo jumlah *followernya* mencapai 1 juta *netizen*. Sementara itu, pasangan calon Sudirman Said dan wakilnya Ida Fauziyah memiliki *follower* sebanyak 28.675 dan 11.721 *follower*. Jumlah *follower* mempengaruhi banyaknya *tweet* yang muncul dari setiap pasangan calon. Selain itu, dapat diketahui bahwa berdasarkan Tabel 4.5 terdapat perbedaan kecenderungan *netizen* dalam menyebut pasangan calon. Pada pasangan calon Sudirman-Ida, *netizen* yang menyebut pasangan calon Sudirman Said dalam *tweet* mereka, pada saat yang bersamaan juga sekaligus cenderung menyebut akun twitter pasangan calon wakilnya yaitu Ida Fauziyah. Sebaliknya pada pasangan calon Ganjar-Yasin *netizen* pada saat menyebut kata akun pasangan calon Ganjar Yasin dalam *tweet* mereka cenderung tidak menyebut akun twitter pasangan calon wakilnya yaiut Taj Yasin. Hal ini menandakan bahwa tingkat popularitas pasangan calon Sudirman-Ida seimbang. Namun tingkat popularitas pasangan calon Ganjar-Yasin berbeda. Dari keseluruhan *tweet netizen* terdapat 2954 jenis unik kata yang muncul dalam *tweet netizen.*

Selanjutnya akan ditampilkan tabel frekuensi relatif data *tweet netizen* terhadap calon kepala daerah Provinsi Jawa Tengah.

Tabel 4.6 Frekuensi Relatif Data *Tweet Netizen* Terhadap Calon Kepala Daerah Provinsi Jawa Tengah

Netizen	Label	Ganjar	Taj Yasin	Sudirman	Ida Fauziyah	Kata 1	Kata 2	...	Kata 2954
1MenujuDamai	2	0,002	0,0	0,008	0,005	0,007	0,006	...	0
4Y4NKZ	1	0,006	0,004	0,001	7.05E+06	0,002	0,002	...	0
AEPishere	1	0,005	0,0	0,001	0,0003	0,002	0,001	...	0
ARIFINjogor	1	0,002	0,072	7.62E+06	7.1E+06	0,001	0,001	...	0
AWPramestuti	1	0,004	0,002	0,0	0,0	0,001	0,001	...	0
Addarul1	2	0,001	0,0	0,005	0,003	0,004	0,004	...	0
Adra94193051	2	0,001	0,0	0,005	0,006	0,006	0,005	...	0
AgungWidrajat	1	0,004	0,0	0,002	0,0004	0,002	0,001	...	0
AhmadTa75651499	1	2.7E+06	0,043	0,0	0,0	0,001	0,0002	...	0
Alim75052441	2	0,0003	0,004	0,005	0,004	0,003	0,007	...	0
AmharMaya	2	0,0001	0,0	0,006	0,004	0,007	0,006	...	0
AnonymWae	2	0,0007	0,0	0,008	0,011	0,006	0,014	...	0
AquinaKinanti	1	0,011	0,018	0,0019	0,0003	0,003	0,004	...	0
Aryprasetyo85	1	0,003	0,0	0,001	7.1E+06	0,002	0,0002	...	0
.		.	.	.	.	.	.	.	.
.		.	.	.	.	.	.	.	.
.		.	.	.	.	.	.	.	.
wongGunung901	2	0.003	0	0.008	0.004	0.011	0.018533	.	0
234		1	1	1	1	1	1	...	1

Sama halnya dengan Provinsi Jawa Timur, cara menghitung frekuensi relatif data *tweet netizen* pada Provinsi Jawa Tengah dilakukan dengan cara menghitung nilai frekuensi setiap unik pada *tweet* masing-masing *netizen*, kemudian dibagi dengan nilai frekuensi kemunculan setiap unik kata dalam keseluruhan *tweet netizen*. Semakin sering *netizen* menyebut kata yang sama dalam *tweet* mereka, maka semakin besar nilai frekuensi relatif kata tersebut. Contoh pada Tabel 4.5 terdapat *netizen* dengan nama akun twitter "AgungWidrajat" yang menyebut kata akun pasangan calon Ganjar Pranowo yaitu "Ganjar" sebanyak 157 kali dari keseluruhan *tweet*nya selama periode kampanye hingga menjelang dilakukan pemungutan suara yaitu pada tanggal 27 Maret hingga 26 Juni 2018. Sedangkan kata akun twitter "Ganjar" selama periode tersebut disebut dalam *tweet* 234 *netizen* berulangkali sebanyak 36.295. Maka selanjutnya untuk menghitung nilai frekuensi relatif *netizen* pada akun twitter "AgungWidrajat" terhadap jenis unik kata "Ganjar" adalah dengan cara membagi nilai frekuensi kemunculan kata "Ganjar", yang dalam hal ini sebagai akun twitter pasangan calon Ganjar Pranowo, pada *tweet netizen* "AgungWidrajat" dengan frekuensi kemunculan kata "Ganjar" dari keseluruhan *tweet netizen* yaitu dalam contoh ini 157 dibagi dengan 36.295. Sehingga dari hasil perhitungan ini diperoleh nilai frekuensi relatif pada Tabel 4.6 yaitu sebesar 0,004. Begitu juga cara untuk untuk mendapatkan nilai frekuensi relatif pada jenis unik kata yang lainnya dari setiap *netizen*. Hasil penjumlahan nilai frekuensi relatif dari setiap unik adalah sama dengan satu. Hasil perhitungan dari nilai frekuensi relatif pada Tabel 4.6 , selanjutnya akan digunakan dalam analisa klasifikasi menggunakan metode CSVM.

Selanjutnya adalah tabel frekuensi data *tweet* terhadap calon kepala daerah Provinsi Jawa Barat dengan jumlah empat pasangan calon.

Tabel 4.7 Frekuensi Data *Tweet Netizen* Terhadap Calon Kepala Daerah Provinsi Jawa Barat

Netizen	Label	Ridwan Kamil	Uu Ruzhanul	Hasanuddin	Anton Charliyan	Sudrajat	Ahmad Syaikhu	Deddy Mizwar	Dedi Mulyadi	Kata 1	Kata 2	...	Kata 2588
2018_bella	3	0	0	0	0	0	0	17	20	12	4	...	0
2019_ganti	3	6	0	6	4	0	0	34	23	27	0	...	0
4626sujono	1	33	3	0	0	0	0	0	0	4	0	...	0
78_challenger	3	3	1	2	2	1	0	25	16	11	0	...	0
ARifqi03267018	1	30	32	0	4	0	0	4	4	2	2	...	0
ARisnawan82	3	6	2	2	3	1	0	28	26	25	2	...	0
ASWHolistik	3	0	0	1	0	0	0	60	53	74	0	...	0
ASYIK_3	3	0	0	0	0	0	0	26	3	9	0	...	0
ATinraga	3	0	0	0	0	0	0	47	15	15	0	...	2
A_Uziii	1	36	13	0	2	0	0	0	0	10	0	...	0
AbdiR31697010	3	0	0	0	0	0	0	25	3	6	0	...	0
.		.	.	.	.	.	.	.	.	.	.	...	.
.		.	.	.	.	.	.	.	.	.	.	...	.
.		.	.	.	.	.	.	.	.	.	.	...	.
zonajjo2	1	37	6	0	0	0	0	0	0	2	0	...	0
562		19515	5539	7699	6293	7699	6293	10710	7917	13113	7122	...	20

Provinsi Jawa Barat diikuti oleh empat pasangan calon pada Pilkada Serentak 2018. Lebih sedikit dibandingkan periode sebelumnya yang dikuti oleh lima pasangan calon dan satu-satunya yang masih bertahan saat ini adalah Dedy Mizwar yang maju sebagai calon gubernur dengan wakilnya Dedi Mulyadi. Sama halnya dengan Provinsi Jawa Tengah, pada Provinsi Jawa Barat hanya sebagian data *tweet* dan jenis unik kata yang digunakan dalam analisis. Tujuannya agar dapat mengurangi beban komputasi saat dilakukan analisis. Terdapat 7181 *tweet*s yang digunakan dalam analisis. *Tweet*s tersebut berasal dari 562 *netizen* menyebut akun twitter salah satu pasangan calon dalam *tweet* mereka. Nama pasangan calon Ridwan Kamil paling sering muncul dalam *tweet netizen* daripada nama pasangan calon lainnya yaitu sebanyak 19.515 kali. Pasangan calon berikutnya yang paling disebut adalah hasanuddin dengan frekuensi muncul sebanyak 13.745 kali. Dedi Mizwar sebagai calon yang diusung oleh Partai Demokrat muncul sebanyak 10.710 kali dalam *tweet netizen*, sedikit lebih banyak dari wakilnya yaitu Dedi Mulyadi yang disebut oleh *netizen* dalam *tweet* mereka sebanyak 7917 kali. Sudrajat sebagai calon nomor urut tiga dan wakilnya Ahmad Syaikhu hanya muncul sebanyak 7669 dan 6293 kali dalam *tweet netizen*. Pasangan calon tersebut merupakan yang paling sedikit muncul dalam *tweet netizen* daripada pasangan calon lainnya. Hal ini disebabkan pasangan calon tersebut lebih cenderung menggunakan kampanye secara langsung daripada kampanye pada media sosial khususnya twitter. Jumlah unik kata dalam keseluruhan *tweet netizen* terhadap empat pasang calon sebanyak 2.588. Kolom label menujukkan preferensi dukungan *netizen* terhadap salah satu pasangan calon yang diperoleh berdasarkan frekuensi menyebut menyebut salah satu pasangan calon. Angka satu menunjukkan bahwa *netizen* lebih sering menyebut pasangan calon nomor urut 1 daripada calon pasangan calon lainnya.

Tabel frekuensi relatif data *tweet* terhadap calon kepala daerah Provinsi Jawa Barat dapat dilihat pada Tabel 4.7.

Tabel 4.8 Frekuensi Relatif Data *Tweet Netizen* Terhadap Calon Kepala Daerah Provinsi Jawa Barat

Netizen	Label	Ridwan Kamil	Uu Ruzhanul	Hasanuddin	Anton Charliyan	Sudrajat	Ahmad Syaikhu	Deddy Mizwar	Dedi Mulyadi	Kata 1	...	Kata 2588
2018_bella	3	0,0	0,0	0,0	0,0	0,0	0,0	0,0015	0,0025	0,0009	...	0
2019_ganti	3	0,0003	0,0	0,0004	0,0003	0,0	0,0	0,0031	0,0029	0,0020	...	0
4626sujono	1	0,0016	0,0005	0,0	0,0	0,0	0,0	0,0	0,0	0,0003	...	0.0
78_challenger	3	0,0001	0,0001	0,0001	0,0001	0,0001	0,0	0,0023	0,0020	0,0008	...	0
ARifqi03267018	1	0,0015	0,0057	0,0	0,0003	0,0	0,0	0,0003	0,0005	0,0001	...	0
ARisnawan82	3	0,0003	0,0003	0,0001	0,0002	0,0001	0,0	0,0026	0,0032	0,0019	...	0
ASWHolistik	3	0,0	0,0	727537	0,0	0,0	0,0	0,0056	0,0066	0,0056	...	0
ASYIK_3	3	0,0	0,0	0,0	0,0	0,0	0,0	0,0024	0,0003	0,0006	...	0.0
ATinraga	3	0,0	0,0	0,0	0,0	0,0	0,0	0,0043	0,0018	0,0011	...	0
A_Uziii	1	0,0018	0,0023	0,0	0,0001	0,0	0,0	0,0	0,0	0,0007	...	0
AbdiR31697010	3	0,0	0,0	0,0	0,0	0,0	0,0	0,0023	0,0003	0,0004	...	0.1
.		.	.	.	.	.	.	.	.	.	...	.
.		.	.	.	.	.	.	.	.	.	...	.
.		.	.	.	.	.	.	.	.	.	...	.
zonajjo2	1	0,001	0,001	0,0	0,0	0,0	0,0	0,0	0,0	0,0001	...	0,0
562		1	1	1	1	1	1	1	1	1	...	1

4.2 Hasil Persebaran *Tweet* .

Hasil praposes data *tweet netizen* akan dijabarkan persebarannya berdasarkan periode dan wilayah yaitu periode kampanye dan hari tenang. Sedangkan untuk wilayah terdiri dari wilayah Jawa Timur, Jawa Tengah dan Jawa Barat.

4.2.1 Persebaran *Tweet* Periode Kampanye dan Hari Tenang

Persebaran data *tweet netizen* mengenai pasangan calon kepala daerah di Jawa Timur, Jawa Tengah, dan Jawa Barat akan dijabarkan mengunakan visualisasi *word cloud*. Visualisasi data teks menggunakan *word cloud* digunakan untuk mengetahui kata-kata yang paling sering muncul pada *tweet netizen* terhadap calon pasangan kepala daerah dari setiap wilayah di Pulau Jawa. Visualisasi *word cloud* menggunakan software Rstudio. Ukuran *font* pada *word cloud* menunjukkan frekuensi kemunculan kata. Semakin besar ukuran *font* menujukkan bahwa semakin besar frekuensi kemunculan kata tersebut.

Hasil visualisasi *word cloud* digunakan untuk mengetahui respons masyarakat pada twitter terhadap pasangan calon kepala daerah dari setiap wilayah selama periode kampanye politik dan hari tenang. Berikut merupakan *word cloud* data *tweet netizen* pada setiap wilayah yaitu Jawa Timur, Jawa Tengah, dan Jawa Barat.

(a)

(b)

(c)

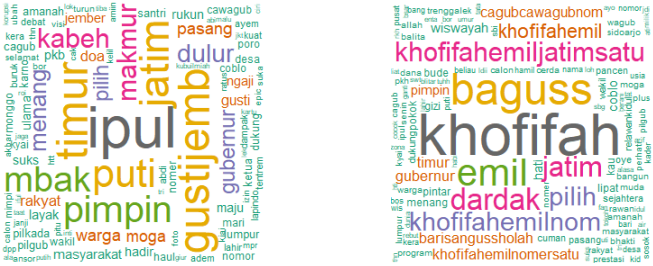
Gambar 4.1 *Word cloud Tweet Netizen* Terhadap Pasangan Calon Kepala Daerah Provinsi Jawa Timur (a), Jawa Tengah (b), dan Jawa Barat (c)

Pasangan calon Khofifah-Emil memiliki frekuensi lebih tinggi diperbincangkan di media sosial twitter dibandingkan lawan politiknya yaitu Gus Ipul-Puti yang melaju pada pilkada serentak 2018 di Provinsi Jawa Timur. Hal ini terlihat pada Gambar 4.1(a) kata "khofifah" dan "emil" memiliki ukuran lebih besar dibandingkan dengan lawan pasangannya yaitu Gus Ipul-Puti. Pada Provinsi Jawa Tengah tampak nama pasangan calon Sudirman Said sedikit lebih tinggi frekuensi munculnya dalam *tweet netizen* dari nama pasangan calon Ganjar Pranowo. Kata "Jateng" mendominasi kemunculannya dalam *tweet netizen* pada provinsi Jawa Tengah. Sedangkan pada wilayah Provinsi Jawa Barat kata "hasanah" lebih sering muncul dalam *tweet netizen* di bandingkan kata lainnya yang mencerminkan nama setiap pasangan calon. Selain nama pasangan calon, juga terdapat kata lain yang sering muncul pada *word cloud* di tiap provinsi misalnya kata "wiswayahe" pada Provinsi Jawa Timur, kata "korupsi" pada Provinsi Jawa Tengah dan kata "teteh" pada Provinsi Jawa Barat.

4.3 Karakteristik Profil Calon Kepala Daerah Berdasarkan Data *Tweet*

Perbincangan politik di media sosial twitter dapat dapat memberikan gambaran kepada publik melalui *tweet netizen* terkait respons masyarakat terhadap calon pasangan yang melaju pada Pilkada Serentak 2018 di setiap provinsi khususnya di Pulau Jawa. Kumpulan kata dari *tweet netizen* mengenai calon pasangan kepala

daerah dapat mencerminkan karakteristik setiap pasangan calon di mata masyarakat. Berikut ini adalah karakteristik profil pasangan calon kepala daerah di Provinsi Jawa Timur, Jawa Tengah, dan Jawa Barat berdasarkan *tweet netizen* dengan menggunakan visualisasi *word cloud*.



(a) (b)

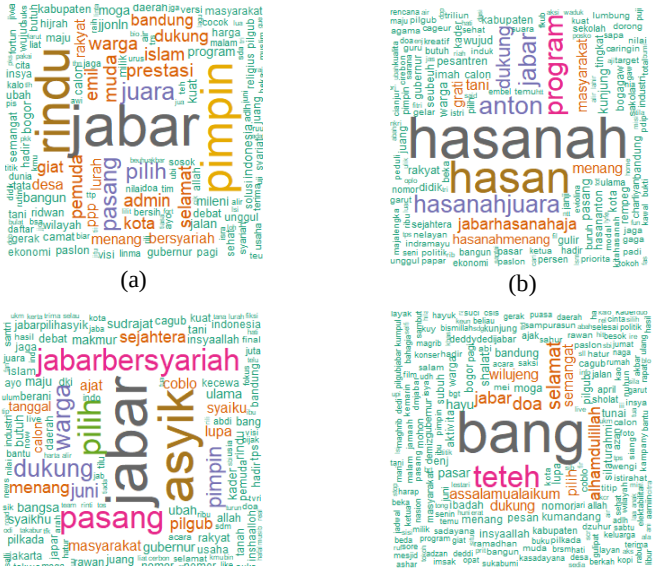
Gambar 4.2 *Word cloud* Pasangan Calon Kepala Daerah Provinsi Jawa Timur Gus Ipul-Puti (a)dan Khofifa-Emil (b)

Word cloud data *tweet netizen* terhadap pasangan Gus Ipul-Puti dan Khofifa-Emil pada Gambar 4.2 menunjukkan kata-kata yang sering muncul pada *tweet netizen* berkaitan dengan kedua calon pasangan kepala daerah tersebut. Kata yang sering muncul berkaitan dengan calon pasangan Gus Ipul dan Puti adalah "pimpin", "ipul", "makmur"dan "gustijemb" dan kata-kata lain yang menjukkan bahwa *netizen* banyak berkeyakinan atau berharap pasangan calon Gus Ipul dan Puti akan membawa Jawa Timur menjadi makmur, adem dan tentram. Begitu juga untuk pasangan calon Khofifa-Emil, *netizen* sering menuliskan kata nama kedua tokoh tersebut dalam *tweet* mereka. Misalnya, kata "khofifah", "emil", dan "baguss" dan kata-kata yang berkaitan dengan nama kedua tokoh tersebut seperti kata "pintar", "cerdas", dan "prestasi".

(a) (b)

Gambar 4.3 *Word cloud* Pasangan Calon Kepala Daerah Provinsi Jawa Tengah Ganjar Pranowo-Taj Yasin (a) dan Sudirman Said - Ida Fauziah (b)

Word cloud pasangan calon kepala daerah Provinsi Jawa Tengah pada Gambar 4.3 terlihat memiliki ukuran *font* yang berbeda antara calon pasangan Ganjar Pranowo-Taj Yasin dan Sudirman Said-Ida Fauziyah. Kata yang sering muncul pada pasangan Ganjar Pranowo–Taj Yasin adalah kata "ganjar", "jateng", dan "gubernur" yang menunjukkan *netizen* mendukung pasangan calon tersebut dalam Pilkada Serentak 2018 di Provinsi Jawa Tengah. Kata-kata lain yang berukuran kecil seperti "menang" dan "warga" menujukkan bahwa frekuensi kemunculannya dalam *tweet netizen* rendah. Sedangkan pada pasangan calon Sudirman Said-Ida Fauziyah kata yang paling sering muncul "jateng", "pimpin", dan "sudirman". Hal ini menandakan ketiga kata tersebut adalah kata yang paling sering muncul dalam *tweet netizen* berkaitan dengan pasangan calon tersebut. Kata-kata lain yang muncul adalah " korupsi" dan " tani". Hal ini berkaitan dengan harapan masyarakat kepada pasangan calon tersebut untuk mengatasi masalah pertanian dan korupsi di Provinsi Jawa Tengah.



Gambar 4.4 *Word cloud* Pasangan Calon Kepala Daerah Provinsi Jawa Barat Ridwal Kamil - Uu Ruzhanul Ulum (a), Tubagus Hasanuddin – Anton Charliyan (b), Sudrajat – Ahmad Syaikhu (c), dan Dedy Mizwar-Dedy Mulyadi (d)

Karakteristik dari pasangan calon Ridwal Kamil-Uu Ruzhanul Ulum berdasarkan data *tweet netizen* pada twitter dapat dilihat pada Gambar 4.4 (a). Banyak *netizen* menilai bahwa calon pasangan tersebut akan membawa Jawa Barat lebih baik. Hal ini dapat dilihat dengan banyaknya kemunculan kata "pimpin", "prestasi", dan "juara" pada *tweet netizen* terhadap pasangan tersebut. Pada pasangan calon Tubagus Hasanuddin–Anton Charliyan kata yang memiliki frekuensi kemunculan paling tinggi dalam *tweet netizen* pada pasangan tersebut adalah "hasanah" dan "hasan". Hal ini erat kaitannya dengan antusias *netizen* pengguna twitter yang mendukung pasangan calon tersebut dalam pilkada nanti. Selain itu juga terdapat kata "tani" pada *word cloud* pasangan tersebut yang berarti bahwa *netizen* menilai pasangan Tubagus Hasanuddin–Anton memiliki kepedulian lebih terhadap para petani. Pada pasangan calon Sudrajat–Ahmad Syaikhu kata

yang mencerminkan karakteristik yang dimiliki pasangan tersebut berdasarkan penilaian masyarakat pada twitter adalah “asyik”. Hal ini dikarenakan “asyik” "singkatan dari nama pasangan calon tersebut. Sedangkan pada pasangan calon Dedy Mizwar-Dedi Mulyadi Kata yang yang sering muncul dalam *tweet netizen* berkaitan dengan pasangan calon tersebut adalah kata “Bang” yang merupakan kata sapaan dari Deddy Mizwar. Kata lain yang sering disebut diantaranya adalah “teteh”, “dukung” dan “doa”.

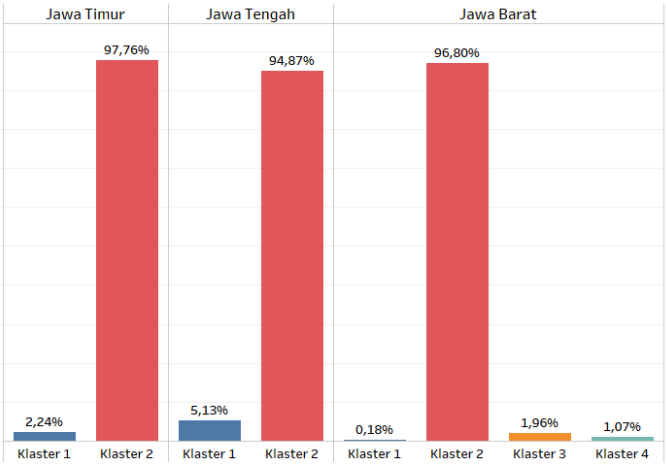
4.4 *Clustered Support Vector Machine Classification*

Algoritma *Clustered Support Vector Machine* pada penelitian ini akan menggunakan algoritma “*kmeans*” sebagai sebagai *clustering* dan dua macam *kernel* sebagai klasifikasi yaitu: *kernel radial basic function* dan *linear*. Pada algoritma *clustering* jumlah klaster ditentukan oleh jumlah pasangan calon dari setiap provinsi yaitu 2 klaster untuk Provinsi Jawa Timur dan Jawa Tengah, dan 4 klaster untuk Provinsi Jawa Barat. Dalam analisis menggunakan metode CSVM, data *tweet netizen* dari setiap provinsi terlebih dahulu akan dikelompokkan ke dalam beberapa klaster, kemudian dari setiap klaster akan dilakukan klasifikasi berdasarkan data *tweet netizen* dengan menggunakan metode SVM. Terdapat dua jenis data *tweet netizen* yang digunakan dalam analisa menggunakan metode *Clustered Support Vector Machine Classification* yaitu data frekuensi dan data frekuensi relatif.

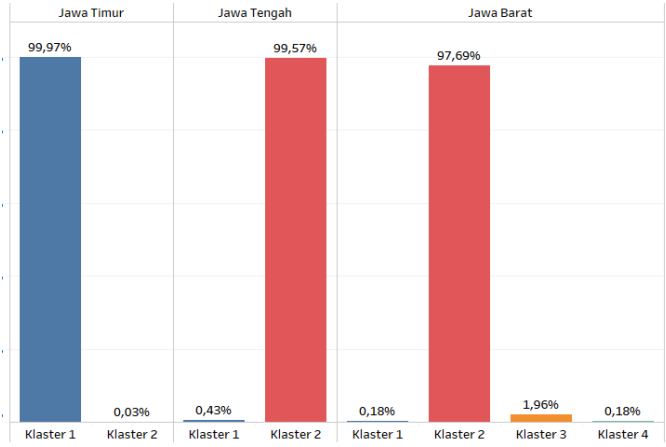
4.4.1 Pengelompokkan *Netizen* Menggunakan Metode K-Means

Pengelompokan *netizen* menggunakan metode *K-Means* pada Provinsi Jawa Timur, Jawa Tengah dan Jawa Barat masing-masing akan dibentuk klaster sejumlah banyaknya pasangan calon kepala daerah dari setiap daerah. 2 klaster untuk provinsi Jawa Timur dan Jawa Tengah dan 4 klaster untuk provinsi Jawa Barat. Hasil analisis klaster akan digunakan untuk mengetahui proporsi pendukung dari setiap pasangan calon dari keseluruhan *netizen* yang melakukan *tweet* berhubungan dengan pasangan calon kepala daerah. Berikut ini adalah persentase proporsi hasil

pengelompokkan *netizen* dengan menggunakan metode *K-Means* pada provinsi Jawa Timur, Jawa Tengah dan Jawa Barat.



(a)

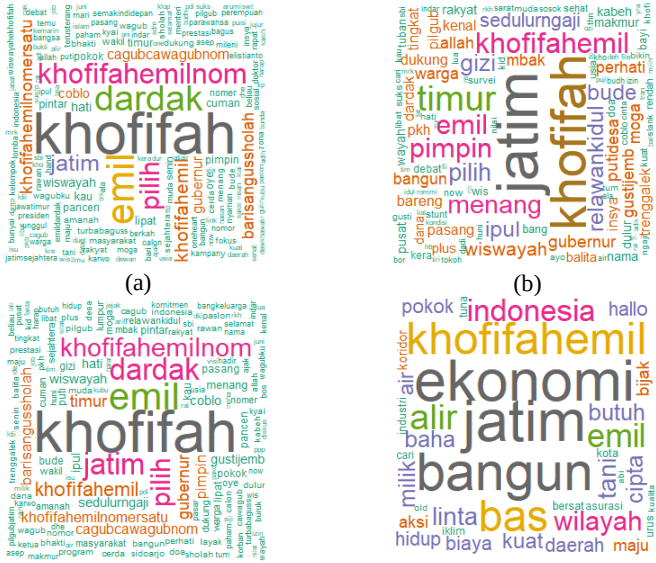


(b)

Gambar 4.5 Proporsi Hasil Klaster *Netizen* Pada Setiap Provinsi berdasarkan Data Frekuensi (a) dan Data Frekuensi Relatif *Tweet Netizen* (b)

Gambar 4.5 memberikan informasi proporsi *netizen* dalam tiap klaster di Provinsi Jawa Timur, Jawa Tengah dan Jawa Barat

berdasarkan data frekuensi (a) dan frekuensi relatif (b) *tweet netizen*. Pada Provinsi Jawa Timur jumlah *netizen* terbagi menjadi dua klaster yaitu klaster 1 dan klaster 2. Hal ini dikarenakan pada provinsi Jawa Timur terdapat dua pasangan calon yang melaju pada Pilkada Serentak 2018. Berdasarkan data frekuensi *tweet netizen* persentase jumlah *netizen* yang masuk dalam klaster 1 sebesar 2,24% dan 97,76% pada klaster 2. Pada Provinsi Jawa Tengah persentase jumlah *netizen* pada klaster 1 sebesar 5,13%, sedangkan pada klaster 2 sebesar 94,87%. Begitu juga pada Provinsi Jawa Barat *netizen* cenderung mengelompok pada salah satu klaster yaitu klaster 2 sebesar 96,80%. Terdapat perbedaan cukup signifikan proporsi netizen pada pada Provinsi Jawa Timur berdasarkan data frekuensi dan frekuensi relatif *tweet netizen*. Berdasarkan data frekuensi, *netizen* cenderung mengelompok pada klaster 2 yakni sebesar 97,76% dan sisanya mengelompok pada klaster 1. Sedangkan jika berdasarkan data frekuensi relatif *tweet netizen* proporsi jumlah *netizen* pada klaster 1 lebih besar dari pada klaster 2 yakni sebesar 99,97%. Hal ini terjadi dikarenakan proses dalam menentukan 2 *centroid* awal dalam analisis klaster menggunakan metode *k-means* dilakukan secara acak sehingga mengakibatkan terjadi perbedaan letak klaster yang lebih besar jumlah anggotanya dari dua jenis data tersebut. Namun selisih proporsinya relatif sama. Sedangkan pada dua provinsi lainnya yaitu Jawa Tengah dan Jawa Barat proporsi *netizen* untuk dua jenis data tersebut relatif sama. Kemudian selanjutnya akan ditampilkan *word cloud* data *tweet netizen* dari masing-masing klaster pada setiap provinsi berdasarkan data frekuensi dan data frekuensi relatif *tweet netizen*. Visualisasi *word cloud* akan memberikan informasi karakteristik *tweet netizen* dari setiap klaster sebagai kelompok pendukung dari salah satu pasangan calon kepala daerah.



Gambar 4.6 *Word cloud* Klaster 1 (a, c) dan Klaster 2 (b,d) Berdasarkan Data Frekuensi (a,b) dan Data Frekuensi Relatif (c,d) *Tweet Netizen* Provinsi Jawa Timur

Gambar 4.6 (a) dan (b) memberikan informasi jenis kata yang sering muncul pada *tweet netizen* pada klaster 1 dan klaster 2 berdasarkan data frekuensi *tweet netizen*. Kata "khofifah" sebagai indikasi pasangan calon Khofifah-Emil mendominasi baik pada klaster 1 maupun klaster 2. Sedangkan lawan politiknya Gus Ipul-Puti yang diindikasikan dengan kata "mbak" dan "ipul" terlihat pada klaster 2 dengan ukuran *font* jauh lebih kecil daripada kata "khofifah". Hal ini menunjukkan bahwa *netizen* lebih sering menyebut pasangan calon Khofifah-Emil daripada Gus Ipul-Puti pada *tweet* mereka. Namun meskipun kata "khofifah" memiliki frekuensi paling tinggi pada klaster 1 dan 2, kata tersebut tidak menjadikan karakteristik pembeda antara klaster 1 dan klaster 2. Karakteristik kata yang muncul pada klaster 1 cenderung lebih berhubungan dengan sifat personal pasangan calon seperti kata "pintar", "amanah" dan "berkah". Sedangkan pada klaster 2

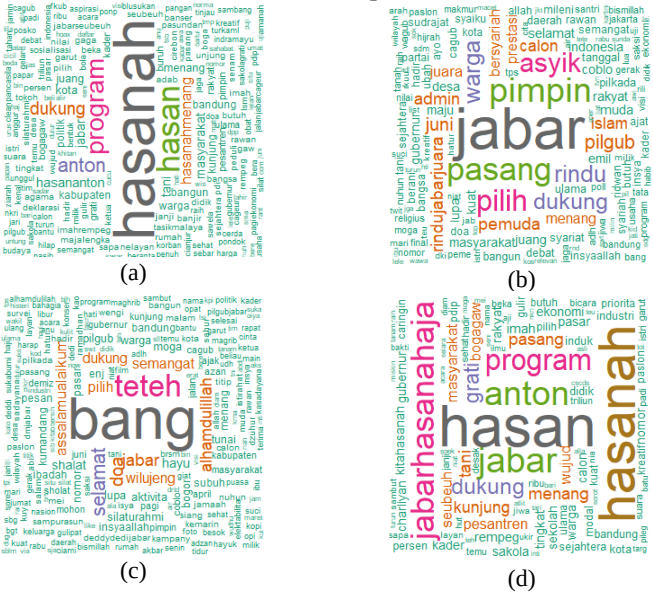
karakteris kata yang muncul cenderung lebih berhubungan dengan program kerja misalnya "pkh", "bangun", dan "makmur". Begitu juga pada visualisasi *word cloud* berdasarkan data frekuensi relatif *tweet netizen* yang menujukkan kata yang berhubungan dengan pasangan calon Khofifah-Emil mendominasi kemunculannya pada klaster 1 dan klaster 2 seperti kata "khofifah" pada klaster 1 dan kata "khofifahemil" pada klaster 2. Sedangkan kata yang berhubungan dengan pasangan calon Gus Ipul-Puti muncul yakni kata "ipul" dan "puti" muncul pada klaster 1 dengan ukuran *font* jauh lebih kecil daripada kata "khofifah" yang mengindikasikan bahwa *netizen* menyebut pasangan tokoh tersebut secara bersamaan dalam *tweet* mereka. Nama provinsi yaitu "jatim" muncul pada klaster 1 dan klaster 2 dengan ukuran *font* lebih kecil daripada kata "khofifah" pada klaster 1 dan paling besar pada klaster 2. Namun kata ini tidak menjadi variabel pembeda dalam antara klaster 1 dan 2 meskipun memiliki frekuensi muncul dalam *tweet netizen* lebih tinggi daripada kata yang lain. Hal ini dikarenakan terdapat banyak kata lain yang mempengaruhi *netizen* mengelompok dalam satu klaster. Variabel kata yang membedakan klaster 1 dan klaster 2 adalah jenis kata yang muncul pada *tweet netizen*. Pada klaster 1 kata yang sering muncul pada *tweet netizen* lebih berhubungan dengan sifat personal dan pilkada serentak 2018 misalnya kata "pilih", "gubernur", dan "pimpin". Sedangkan pada klaster 2 kata yang muncul cenderung lebih berhubungan dengan program pemerintah misalnya kata "daerah", "maju", dan "biaya". Sehingga berdasarkan hasil visualisasi dari dua jenis data tersebut dapat dikatakan bahwa minat masyarakat terhadap pasangan calon Khofifah-Emil lebih tinggi daripada pasangan calon Gus Ipul-Puti. Hasil visualisasi ini juga sejalan dengan fakta yang ada di lapangan pada saat pelaksanaan pemilu berdasarkan hasil *real count* KPUD Jawa Timur menunjukkan bahwa pasangan Khofifa-Emil memperoleh suara lebih banyak dibandingkan pasangan Gus Ipul-Puti yaitu sebesar 53,73%. Sedangkan pasangan Gus Ipul-Puti mendapatkan perolehan suara sebesar 46,27% (Flora, 2018). *Word*

cloud klaster data *tweet netizen* pada Provinsi Jawa Tengah dapat dilihat pada Gambar 4.7.

Gambar 4.7 *Word cloud* Klaster 1 (a, c) dan Klaster 2 (b,d) Berdasarkan Data Frekuensi (a,b) dan Data Frekuensi Relatif (c,d) *Tweet Netizen* Provinsi Jawa Tengah

Visualisasi *word cloud* pada Gambar 4.7 menjelaskan kata yang sering muncul pada *tweet netizen* berkaitan dengan calon pasangan kepala daerah Pronvisi Jawa Tengah. Karakteristik yang tampak pada Gambar 4.7 adalah klaster 1 baik berdasarkan kata frekuensi maupun frekuensi relatif *tweet netizen* didominasi oleh nama pasangan calon Sudirman Said. Sedangkan pada klaster 2 didominasi oleh kata nama pasangan calon Ganjar Pranowo, meskipun juga terdapat kata “sudirman” dan “said” dengan ukuran *font* relatif lebih kecil. Nama provinsi yaitu kata “Jateng” muncul dengan ukuran *font* paling besar dibandingkan kata lain di setiap klaster dan jenis data. Hal ini menunjukkan bahwa *netizen* lebih sering menyebut nama provinsi daripada nama pasangan calon meskipun kata ini tidak menjadi variabel pembeda antar klaster di setiap jenis data. Nama pasangan calon menjadi variabel pembeda

antar klaster 1 dan klaster 2. Karakteristik kata yang muncul pada klaster 1 di setiap jenis data lebih berhubungan sifat personal pasangan calon Sudirman Said-Ida Fauziyah seperti kata "sudirman" dan "said". Sedangkan pada klaster 2 lebih dominan berhubungan dengan sifat personal pasangan calon Ganjar Pranowo-Taj Yasin misalnya kata "ganjaryasin" dan "ganjar". Hasil ini sedikit bertentangan dengan kondisi politik di dunia nyata. Berdasarkan hasil rekapitulasi penghitungan suara oleh KPU Jawa Tengah, pasangan dengan nomor urut 1 tersebut menang cukup banyak dari lawan politiknya Sudirman Said-Ida Fauziyah dengan persentase total perolehan suara pasangan nomor 1 sebesar 58,78% dan 41,22% untuk pasangan nomor urut 2 (Purbaya, 2018). Sedangkan karakteristik kata yang sering muncul pada *tweet netizen* pada Provinsi Jawa Barat berdasarkan data frekuensi *tweet netizen* dapat dilihat pada Gambar 4.8.



Gambar 4.8 *Word cloud* Klaster 1(a), Klaster 2(b), Klaster 3(c), dan Klaster 4(d) Berdasarkan Data Frekuensi *Tweet Netizen* Provinsi Jawa Barat

Hasil visualisasi *word cloud* klaster Provinsi Jawa Barat berdasarkan data frekuensi *tweet netizen* terlihat pada Gambar 4.8. Berdasarkan gambar tersebut dapat dilihat bahwa pada klaster 1 dan 4 didominasi oleh kata "hasanah". Pada klaster 2 nama provinsi yaitu kata "jabar" muncul sebagai kata yang mempunyai frekuensi paling besar disebut oleh *netizen*. Sedangkan pada klaster 3 kata "bang" sebagai indikasi pasangan calon Dedy Mizwar muncul dengan ukurang *font* paling besar daripada kata lainnya yang menujukkan bahwa kata ini muncul lebih sering daripada kata lainnya dalam *tweet netizen* pada klaster 3. Karakteristik kata yang membedakan antar klaster adalah jenis kata yang muncul pada setiap klaster. Pada klaster 1 karakteritik kata yang muncul cenderung berhubungan dengan sifat personal pasangan calon Tubagus Hasanuddin-Anton Charliyan dari akun pribadi *netizen* seperti kata "semangat", "juang", dan "aspirasi". Pada klaster 2 karakteristik kata yang tampak membedakan dengan klaster lain adalah adanya kata "asyik" dengan ukuran cukup besar yang mengindikasikan dari nama pasangan calon Sudrajat-Ahmad Syaikhu. Sehingga dapat dikatakan pada klaster 3 karakteristik kata yang muncul cenderung berhubungan dengan pasangan calon Sudrajat-Ahmad Syaikhu. Kata "Bang" pada klaster 3 mengindikasikan bahwa pada klaster tersebut kata yang muncul cenderung berhubungan dengan pasangan calon Dedy Mizwar-Dedy Mulyadi. Sedangkan pada klaster 4 hampir sama dengan klaster 1 namun yang membedakan adalah pada klaster 4 kata yang muncul kebanyakan berasal dari akun twitter tim sukses pasangan calon Tubagus Hasanuddin-Anton Charliyan. Hal ini dapat dibuktikan dengan adanya kata "jabarhasanahaja" dengan ukuran *font* cukup besar pada *word cloud*. Kata ini bersumber dari akun twitter tim sukes pasangan calon Tubagus Hasanuddin-Anton Charliyan yaitu jabar hasanah. Sehingga kata yang muncul pada klaster 4 cenderung berhubungan kata kegiatan kampanye baik pada dunia *real* maupun dunia maya misalnya kata "dukung", "pesantren" dan "kunjung". Selanjutnya adalah gambar visualisasi

word cloud berdasarkan data frekuensi relatif *tweet netizen* pada Provinsi Jawa Barat.

(a) (b) (c) (d)

Gambar 4.9 *Word cloud* Klaster 1(a), Klaster 2(b), Klaster 3(c), dan Klaster 4(d) Berdasarkan Data Frekuensi Relatif *Tweet Netizen* Provinsi Jawa Barat

Hasil visualisasi *word cloud* menggunakan data frekuensi relatif *tweet netizen* tidak jauh berbeda dengan menggunakan data frekuensi *tweet netizen*. Hanya saja terdapat perbedaan letak klaster dari visualisasi *word cloud*. Gambar visualisasi *word cloud* klaster 4 pada Gambar 4.8 karakteristik kata yang muncul relatif sama dengan gambar visualisasi *word cloud* klaster 2 pada gambar 4.9. Pada klaster 4 kata "hasanah" dan "hasan" memiliki ukuran *font* lebih besar daripada kata lainnya. Hal ini menunjukkan bahwa *netizen* lebih sering menyebut kata tersebut pada *tweet* mereka yang mengindikasikan nama pasangan calon Tubagus Hasanuddin-Anton Anton Charliyan. Sedangkan pada klaster 1 dan 3 karakteristik kata yang muncul sama halnya dengan pada klaster 1 dan 3 pada Gambar 4.8. Hasil visualisasi *word cloud* pada Gambar 4.8 dan 4.9 cukup berbeda jauh dengan fakta yang ada di

lapangan. Hasil perhitungan suara yang dilakukan oleh KPU Jawa Barat menunjukkan bahwa pasangan Ridwan Kamil-Uu Ruzhanul Ulum memperoleh suara sebesar 32,88%, Tubagus Hasanuddin-Anton Charliyan 12,62%, Sudrajat-Ahmad Syaikhu 28,74% dan pasangan calon Dedy Mizwar-Dedi Mulyadi memperoleh suara sebesar 25,77% (Ramdhani, 2018).

4.4.2 CSVM Menggunakan *Kernel* Linear

Pembahasan klasifikasi data *tweet netizen* menggunakan metode CSVM *kernel linear* akan mempertimbangkan nilai parameter C untuk mendapatkan ketepatan klasifikasi terbaik. Nilai parameter C yang akan digunaan rentang nilai 10^{-3} hingga 10^{3}. Model yang diperoleh dari setiap klaster selanjutnya akan diperoleh untuk mengklasifikasikan data *testing* dari setiap klaster. Hasil ketepatan klasifikasi dari pada data *testing* dari setiap provinsi dapat dilihat pada Lampiran 1 dan Lampiran 2. Berikut merupakan hasil ketepatan klasifikasi dari setiap provinsi menggunakan klasifikasi CSVM *kernel linear*.

Tabel 4.9 Ketepatan Klasifikasi CSVM *Kernel Linear*

Data	Provinsi	C	Akurasi	G-Mean	AUC	*Time*
Frekuensi	Jatim	1	1	1	1	7,157
	Jateng	1	0,978	0,970	0,971	1,731
	Jabar	1	0,987	0,885	0,892	2,329
Frekuensi Relatif	Jatim	100	0,945	0,920	0,921	31,226
	Jateng	1	0,971	0,960	0,961	3,39
	Jabar	1	0,930	0,666	0,712	4,609

Ketepatan klasifikasi terbaik menggunakan CSVM *kernel linear* pada Tabel 4.9 diperoleh dari hasil perhitungan menggunakan parameter C. Pada Provinsi Jawa Timur dengan menggunakan data frekuensi *tweet netizen* tingkat ketepatan klasifikasinya sebesar 100% dengan nilai parameter C sebesar 1. Ketepatan klasifkasi terbaik 97,8% pada data frekuensi *tweet netizen* terhadap pasangan calon kepala daerah Provinsi Jawa Tengah diperoleh dari nilai parameter C sebesar 1. Selain itu, nilai parameter tersebut memberikan hasil *G-Mean* dan AUC berturut-

turut sebesar 97% dan 97,1%. Begitu juga pada Provinsi Jawa Barat nilai ketepatan klasifikasi terbaik 98,7% diperoleh dengan menggunakan parameter C sebesar 1. Sedangkan hasil ketepatan klasifikasi terbaik menggunakan data frekuensi relatif *tweet netizen* pada Provinsi Jawa Timur 94,5% dengan nilai parameter C sebesar 100. Dua provinsi lainnya yakni Jawa Tengah dan Jawa Barat diperoleh nilai akurasi sebesar 97,1% dan 93% dari hasil optimasi menggunakan prameter C sebesar 1 dengan waktu komputasi selama 3,39 detik dan 4,609 detik. Secara keseluruhan hasil klasifikasi metode *kernel linear* dengan menggunakan data frekuensi *tweet netizen* dapat memberikan hasil nilai akurasi lebih tinggi dibandingkan menggunakan data frekuensi relatif *tweet netizen* pada Provinsi Jawa Timur, Jawa Tengah dan Jawa Barat.

4.4.3 CSVM Menggunakan *Kernel* RBF

Pada klasifikasi CSVM dengan *kernel* RBF terdapat 2 parameter yang digunakan yaitu parameter *C* dan parameter gamma. Nilai parameter C dan parameter gamma yang digunakan adalah rentang nilai 10^{-2} hingga 10^{2}. Hasil perhitungan ketepatan klasifikasi dan waktu komputasi dari setiap provinsi dengan menggunakan CSVM *kernel* RBF dapat dilihat pada Lampiran 3 hingga Lampiran 8. Hasil ketepatan klasifikasi terbaik dan waktu komputasi pada Provinsi Jawa Timur, Jawa Tengah dan Jawa Barat dengan menggunakan metode CSVM untuk *kernel* RBF dapat dilihat pada Tabel 4.10 berikut.

Tabel 4.10 Ketepatan Klasifikasi CSVM *Kernel* RBF

Data	Provinsi	C	Gamma	Akurasi	G-Mean	AUC	*Time*
Frekuensi	Jatim	1	0,01	0,999	0,999	0,999	13,447
	Jateng	0,01	0,01	0,938	0,921	0,925	2,854
	Jabar	0,1	0,01	0,973	0,812	0,830	6,405
Frekuensi Relatif	Jatim	100	100	0,929	0,891	0,894	34,029
	Jateng	10	1	0,952	0,935	0,937	2,67
	Jabar	100	0,1	0,956	0,765	0,788	6,054

Ketepatan klasifikasi terbaik menggunakan CSVM *kernel* RBF pada Tabel 4.10 diperoleh dari hasil kombinasi parameter C

dan gamma. Pada Provinsi Jawa Timur dengan menggunakan data frekuensi *tweet netizen* diperoleh tingkat ketepatan klasifikasi terbaik sebesar 99,9%. Nilai tersebut diperoleh dari hasil kombinasi parameter C dan gamma sebesar 1 dan 0,01. Ketepatan klasifikasi terbaik 93,8% pada data frekuensi *tweet netizen* terhadap pasangan calon kepala daerah Provinsi Jawa Tengah diperoleh dari kombinasi nilai parameter C dan gamma sebesar 0,01. Selain itu, kombinasi nilai parameter tersebut memberikan hasil *G-Mean* dan AUC berturut-turut sebesar 92,1% dan 92,5%. Selanjutnya pada Provinsi Jawa Barat dengan menggunakan data yang sama nilai ketepatan klasifikasi terbaik yang diperoleh sebesar 97,3% dengan nilai parameter C sebesar 0,1 dan gamma sebesar 0,01. Sedangkan hasil ketepatan klasifikasi menggunakan data frekuensi relatif pada Provinsi Jawa Timur, Jawa Tengah dan Jawa Barat sebesar 92,9%, 95,2% dan 95,6%. Nilai akurasi tersebut sedikit lebih rendah jika dibandingkan dengan hasil akurasi menggunakan data frekuensi *tweet netizen*. Secara keseluruhan pada tiap provinsi penggunaan data frekuensi *tweet netizen* memberikan nilai akurasi lebih tinggi dibandingkan dengan menggunakan data frekuensi relatif *tweet netizen.* Provinsi Jawa timur memiliki nilai akurasi paling tinggi dari dua provinsi lainya yakni sebesar 99,9%.

Pengunaan metode CSVM *kernel linear* dan RBF dalam klasifikasi data *tweet netizen* terhadap pasangan calon kepala daerah di Provinsi Jawa Timur, Jawa Tengah dan Jawa Barat memberikan hasil ketepatan klasifikasi sangat tinggi dan efisien. Nilai akurasi menggunakan metode CSVM *kernel linear* lebih tinggi dibandingkankan menggunakan metode CSVM *kernel* RBF pada Provinsi Jawa Timur, Jawa Tengah dan Jawa Barat.

4.4.4 Model CSVM *Kernel* Linear

Berdasarkan pembahasan pada sub bab 4.4.2 dan 4.4.3 diperoleh bahwa metode CSVM *kernel linear* memberikan nilai ketepatan klasifikasi *linear* lebih baik dibandingkan metode CSVM *kernel* RBF pada ketiga provinsi yaitu Jawa Timur, Jawa Tengah dan Jawa Barat. Nilai parameter C yang digunakan agar

mendapatkan nilai ketapatan klasifikasi terbaik pada pada Provinsi Jawa Timur, Jawa Tengah dan Jawa Barat adalah sebesar 1. Berikut adalah fungsi *hyperlane* yang dibentuk dari fungsi *kernel* sesuai persamaan pada tabel 2.1 untuk mengklasfikasikan preferensi dukungan terhadap salah satu pasangan calon kepala daerah di tiap provinsi berdasarkan data *tweet netizen.*

Tabel 4.11 Persamaan *Hyperlane* pada Provinsi Jawa Timur, Jawa Tengah, dan Jawa Barat

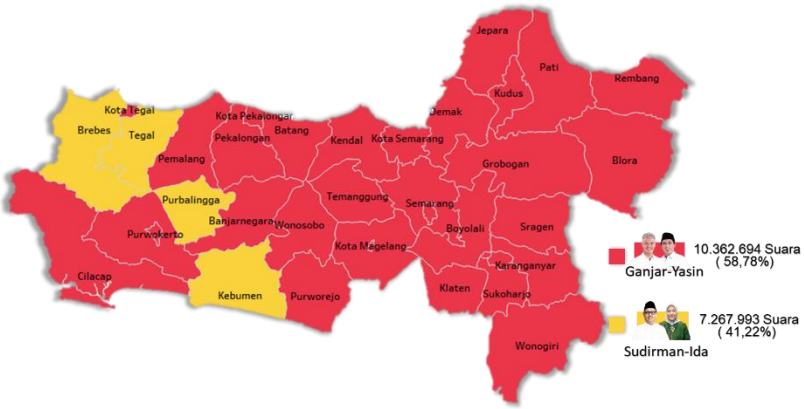
Provinsi	Persamaan *Hyperlane*
Jawa Timur	$f(x) = (\sum_{i=1}^{1728} \alpha_i y_i x_i^T x_j + b)$
Jawa Tengah	$f(x) = (\sum_{i=1}^{2954} \alpha_i y_i x_i^T x_j + b)$
Jawa Barat	$f(x) = (\sum_{i=1}^{2588} \alpha_i y_i x_i^T x_j + b)$

4.5 Peta Perolehan Suara Pilkada Serentak 2018 di Pulau Jawa

Komisi Pemilihan Umum telah mengumumkan hasil rekapitulasi suara dari sejumlah pemilihan kepala daerah di Indonesia. Hasil rekapitulasi suara Pemilihan Gubernur Jawa Timur menunjukkan pasangan Khofifah Indar Parawansa-Emil Elestianto Dardak unggul atas lawannya, Saifullah Yusuf (Gus Ipul)-Puti Guntur. Pasangan Khofifah-Emil memperoleh suara sebesar 10.465.218 , sedangkan lawan pasangannya Gus Ipul-Puti yang memperoleh suara sebesar 9.076.014. Berdasarkan hasil rekapitulasi KPU pasangan Khofifah Indar Parawansa-Emil Elestianto Dardak dinyatakan sebagai pemenang pada pilkada 2018 Provinsi Jawa Timur. Peta kemenangan pasangan Khofifah-Emil dapat dilihat pada Gambar 4.10.

Gambar 4.10 Peta Perolehan Suara Pilkada 2018 Provinsi Jawa Timur

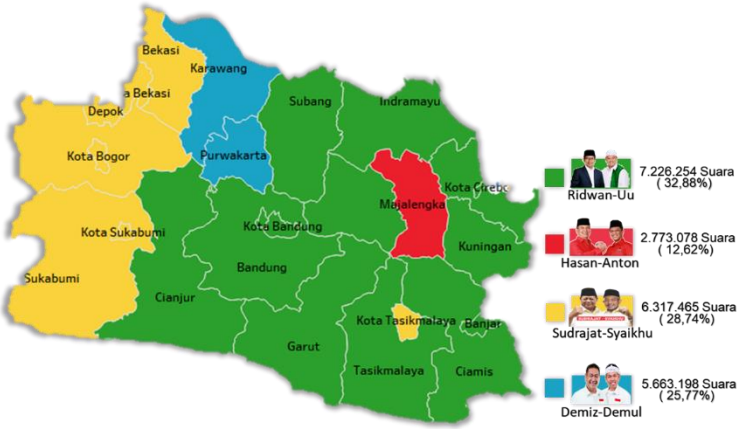
Gambar 4.10 merupakan peta kemenangan hasil pemilihan kepala daerah Provinsi Jawa Timur. Pemilihan kepala daerah Pronvinsi Jawa Timur diikuti oleh 30.155.71 pemilih yang tersebar di 38 kabupaten dan kota dengan jumlah surat suara sah sebanyak 19.541.232 suara dan surat suara tidak sah sebesar 782.027 suara. Pasangan calon Khofifa-Emil unggul atas pasangan Gus Ipul-Puti di 22 Kabupaten dan 5 kabupaten kota. Sedangkan pasangan Gus Ipul-Puti meraup suara lebih tinggi dari lawanya di 7 Kabupaten dan 4 Kabupaten Kota, satu diantaranya adalah tanah kelahirannya yaitu Kabupaten Pasuruan. Wilayah mataraman dan tapal kuda menjadi kunci faktor kemenangan pasangan calon nomor satu. Hanya empat kabupaten pasangan Khofifah-Emil kalah dari pasangan Gus Ipul-Puti yaitu pada Kabupaten Kediri, Kabupaten dan Kota Madiun, Kabupaten Situbondo dan Kabupaten Bangkalan. Sedangkan wilayah kemenangan pasangan nomor urut dua berada pada wilayah arek yaitu meliputi Surabaya hingga Malang. Sehingga berdasarkan gambar peta perolehan suara pilkada 2018 Provinsi Jawa Timur dapat disimpulkan bahwa secara global hasil analisis klaster dapat merefleksikan keadaan *real* politik pada Provinsi Jawa Timur kecuali pada Kabupaten Madiun, Kediri, Blitar, Malang, Situbondo, Bangkalan, Pasuruan, serta Kota Batu. Peta perolehan suara hasil pemilihan kepala daerah Provinsi Jawa Tengah dapat dilihat pada Gambar 4.11.



Gambar 4.11 Peta Perolehan Suara Pilkada 2018 Provinsi Jawa Tengah

Provinsi Jawa Tengah, berdasarkan data KPU, diikuti jumlah pemilih sebanyak 27.068.125 pemilih dengan jumlah surat suara sah sebanyak 17.630.687 dan surat suara tidak sah sebanyak 778.805. Pasangan calon gubernur dan wakil gubernur, Ganjar Pranowo dan Taj Yasin, memenangi Pilkada Jawa Tengah 2018. Hasil rekapitulasi suara KPU Provinsi Jawa Tengah menunjukkan Ganjar-Yasin memperoleh persentase 58,78 persen dengan perolehan 10.362.694 suara. Sementara itu, pasangan Sudirman Said-Ida Fauziyah memperoleh persentase 41,22 persen dengan perolehan 7.267.993 suara. Pasangan Ganjar-Yasin berhasil mengungguli jumlah suara pasangan Sudirman-Ida hampir di seluruh wilayah dari 35 kabupaten dan kota. Pasangan nomor urut satu tersebut kalah dalam jumlah suara dari lawannya Sudirman-Ida di empat kabupaten yaitu Kabupaten Brebes, Tegal, Purbalingga dan Kebumen. Pasangan calon Sudirman Said menang mutlak di wilayah tanah kelahirannya yaitu Kabupaten Brebes dengan persentase perolehan suara sebesar 60,47%. Sedangkan pasangan Ganjar pranowo memperoleh persentase 57,09% persen di wilayah tanah kelahirannya yaitu Kabupaten Karanganyar. Sehingga berdasarkan gambar peta perolehan suara Pilkada 2018 Provinsi Jawa Tengah dapat disimpulkan bahwa secara global hasil analisis klaster tidak dapat merefleksikan

keadaan *real* politik pada Provinsi Jawa Tengah kecuali pada Kabupaten Brebes, Tegal, Purbalingga, serta Kabupaten Kebumen. Selanjutnya akan ditampilkan peta perolehan suara pemilihan kepala daerah Provinsi Jawa Barat pada Gambar 4.12.



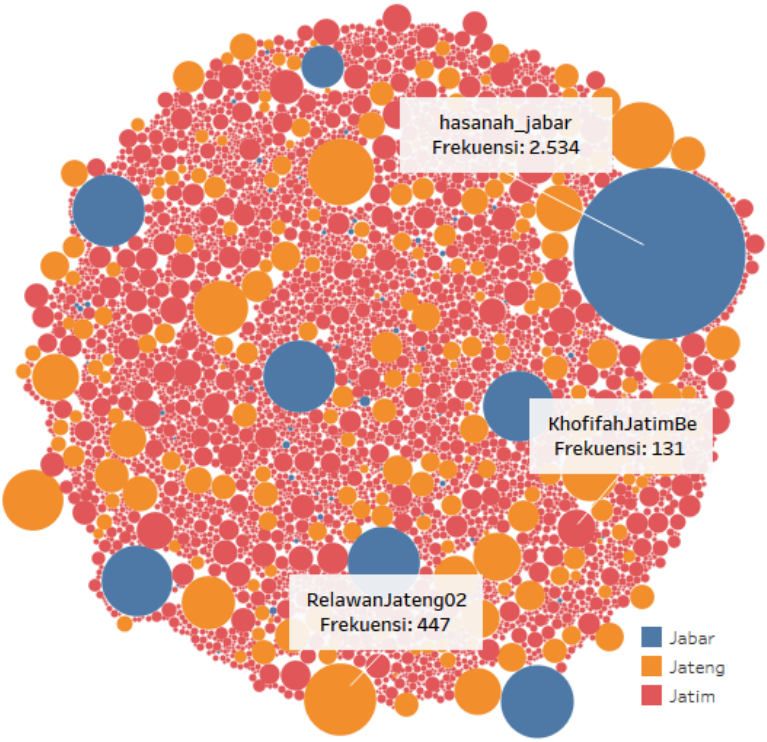
Gambar 4.12 Peta Perolehan Suara Pilkada 2018 Provinsi Jawa Barat

Pada Provinsi Jawa Barat pasangan nomor urut 1, Ridwan Kamil dan Uu Ruzhanul Ulum, berhasil mendapatkan suara terbanyak berdasarkan hasil rekapitulasi dari KPU Provinsi Jawa Barat. Pasangan Ridwan-Uu mendapatkan 7.226.254 suara atau unggul dengan torehan 32,88 persen. Pasangan Ridwan-Uu unggul dengan selisih 4,14 persen dari pesaing terdekatnya, yakni pasangan nomor urut tiga, Sudrajat-Ahmad Syaikhu, dengan raihan 6.317.465 suara atau setara dengan hitungan 28,74 persen. Sedangkan untuk pasangan nomor urut empat, Deddy Mizwar-Dedi Mulyadi, mendapatkan suara 5.663.198 atau 25,77 persen. Terakhir, pasangan nomor urut dua, Tubagus Hasanuddin-Anton Charliyan, mendapatkan suara 2.773.078 atau 12,62 persen. Pasangan Ridwan-Kamil menang di lima belas daerah dari 27 kabupaten dan kota di Jabar. Mayoritas wilayah yang berhasil dikuasai merupakan wilayah pasundan dan pantura seperti

Bandung, Cianjur, Subang, dan Indramayu. Sedangkan pasangan Sudrajat-Syaikhu unggul di delaman daerah yang merupakan wilayah penyanggah ibu kota seperti Bekasi, Bogor, dan Depok. Sementara itu, pasangan Demiz-Dedmul hanya menang di empat daerah, yaitu Purwakarta, Karawang, Subang (Purwasuka) dan Cianjur. Tb-Anton Charliyan menang di dua daerah yaitu Majalengka dan Pangandaran. Sehingga berdasarkan gambar peta perolehan suara Pilkada 2018 Provinsi Jawa Barat dapat disimpulkan bahwa secara global hasil analisis klaster tidak dapat merefleksikan keadaan politik pada Provinsi Jawa Barat kecuali pada Kabupaten Majalengka.

4.6 *Bubble Charts*

Bubble Charts merupakan variasi dari grafik pencar atau *scatterplot* di mana titik data diganti dengan gelembung, dan dimensi tambahan data diwakili dalam ukuran gelembung. Hasil visualisasi *word cloud tweets netizen* pada sub bab 4.2.1 memberikan informasi jenis kata yang paling sering muncul pada *tweets netizen* terhadap pasangan calon kepala daerah di setiap Provinsi. Pada Provinsi Jawa Timur kata yang paling sering muncul adalah kata "Khofifah", dan pada Provinsi Jawa Tengah adalah kata "Jateng". Sementara itu, kata "Hasanah" merupakan kata yang sering muncul pada *tweets netizen* berkaitan dengan pasangan calon kepala daerah Provinsi Jawa Barat. *Bubble Charts* digunakan untuk mengetahui visualisasi sebaran frekuensi *tweet* kata populer pada masing-masing provinsi. Hasil visualisasi *Bubble Charts* akan memberikan informasi sebaran *netizen* dengan jumlah frekuensi *tweets* yang yang mengandung kata populer. Tingkat frekuensi akan mempengaruhi besaran ukuran gelembung. Semakin besar nilai frekuensi *tweets* yang mengandung kata populer, maka semakin besar ukuran gelembung pada *Bubble Charts*. Visualisasi *Buble Charts tweets netizen* terhadap kata populer pada masing-masing provinsi yakni Jawa Timur, Jawa Tengah dan Jawa Barat dapat dilihat pada Gambar 4.13.



Gambar 4.13 *Bubble Charts Tweets Netizen* Terhadap Kata Populer Pada Provinsi Jawa Timur, Jawa Tengah dan Jawa Barat

Pada Provinsi Jawa Timur terlihat pada Gambar 4.13 bahwa *netizen* cenderung memiliki frekuensi yang sama dalam menyebut kata "Khofifah" dalam *tweets* mereka. Hal ini dibuktikan dengan ukuran gelembung warna merah cenderung memiliki ukuran sama. Sementara itu, pada Provinsi Jawa Tengah dan Jawa Barat berdasarkan visualisasi *Bubble Charts* pada Gambar 4.13 dapat dikatakan bahwa kemunculan kata populer pada kedua provinsi ini yaitu kata "Jateng" dan kata "hasanah" berasal dari beberapa akun tertentu saja. Pada Provinsi Jawa Tengah *netizen* yang memiliki frekuensi paling tinggi dalam menyebut kata "jateng" adalah akun

twitter “RelawanJateng02” dengan frekuensi 447 *tweets* yang diduga sebagai akun twitter tim sukses pasangan calon Ganjar Pranowo jika dilihat berdasarkan hasil *crawling* data *tweets*. Sedangkan pada Provinsi Jawa Barat kemunculan kata populer “hasanah” didominasi oleh *tweets* akun *twitter* hasanah_jabar dengan frekuensi 2534 *tweets*. Akun ini diduga merupakan akun twitter tim sukses pasangan calon Tubagus Hasanuddin-Anton Charliyan. Hasil visualisasi *Bubble Charts*, jika dihubungkan dengan hasil analisis klaster menunjukkan adanya indikasi faktor penyebab hasil analisis klaster pada Provinsi Jawa Tengah dan Jawa Barat berbeda dengan hasil dilapangan. Faktor penyebab hasil analisis klaster pada provinsi Jawa Tengah dan Jawa Barat berbeda dengan hasil di lapangan adalah diduga karena adanya dominasi *tweet* dari beberapa *netizen* terhadap pasangan calon pada kedua provinsi tersebut. Sehingga berdasarkan hasil visualisasi *Bubble Charts* dapat disimpulkan bahwa banyaknya *tweet netizen* tidak dapat merefleksikan keadaan politik di lapangan, tapi dapat direfleksikan oleh banyaknya *netizen* yang *tweet* di twitter.

(*Halaman ini sengaja dikosongkan*)

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Kesimpulan yang diperoleh dari hasil analisis penelitian ini adalah sebagai berikut.

1. Kata yang sering muncul pada Provinsi Jawa Timur adalah seputar pasangan calon Khofifa-Emil. Pada Provinsi Jawa Tengah didominasi kata "jateng" dan "sudirman". Sedangkan pada Provinsi Jawa Barat kata yang sering muncul pada *tweet netizen* adalah "hasanah", "bang" dan "program".
2. Hasil pengelompokkan *netizen* berdasarkan data frekuensi *tweet netizen* memenggunakan metode *k-means* pada Provinsi Jawa Timur adalah terdapat 2,24% *netizen* berada pada klaster 1 dan 97,76% pada klasters 2. Pada Provinsi Jawa Tengah 5,13% *netizen* berada padala klaster 1 dan 94,87% pada klaster 2. Sedangkan pada Provinsi Jawa Barat 0,18% *netizen* berada pada klaster 1, 96,80% berada pada klaster 2, 1,96% pada klaster 3 dan sisanya 1,07% *netizen* berada pada klaster 4.
3. Hasil pengelompokkan *netizen* berdasarkan data frekuensi relatif *tweet netizen* memenggunakan metode *k-means* pada Provinsi Jawa Timur adalah terdapat 99,97% *netizen* berada pada klaster 1 dan 0,03% pada klasters 2. Pada Provinsi Jawa Tengah 0,43% *netizen* berada padala klaster 1 dan 99,57% pada klaster 2. Sedangkan pada Provinsi Jawa Barat 0,18% *netizen* berada pada klaster 1, 97,69% berada pada klaster 2, 1,96% pada klaster 3 dan sisanya 0,18% *netizen* berada pada klaster 4.
4. Hasil ketepatan klasifikasi menggunakan metode CSVM *kernel linear* berdasarkan data frekuensi *tweet netizen* pada Provinsi Jawa Timur, Jawa Tengah dan Jawa Barat secara berturut-turut adalah sebesar 100%, 97,8% dan 98,7%. Waktu komputasi yang dibutuhkan dalam proses running CSVM *kernel linear* pada tiga provinsi tersebut adalah 7,157 detik 1,731 detik dan 2,329 detik.

5. Hasil ketepatan klasifikasi menggunakan metode CSVM *kernel linear* berdasarkan data frekuensi relatif *tweet netizen* pada Provinsi Jawa Timur, Jawa Tengah dan Jawa Barat secara berturut-turut adalah sebesar 94,5%, 97,1% dan 93%. Waktu komputasi yang dibutuhkan dalam proses running CSVM *kernel linear* pada tiga provinsi tersebut adalah 31,226 detik 3,39 detik dan 4,609 detik.
6. Hasil ketepatan klasifikasi menggunakan metode CSVM *kernel* RBF berdasarkan data frekuensi *tweet netizen* pada Provinsi Jawa Timur, Jawa Tengah dan Jawa Barat secara berturut-turut adalah sebesar 99,9%, 93,8% dan 97,3%. Waktu komputasi yang dibutuhkan dalam proses running CSVM *kernel linear* pada tiga provinsi tersebut adalah 13,44 detik, 2,854 detik dan 6,40 detik.
7. Hasil ketepatan klasifikasi menggunakan metode CSVM *kernel* RBF berdasarkan data frekuensi relatif *tweet netizen* pada Provinsi Jawa Timur, Jawa Tengah dan Jawa Barat secara berturut-turut adalah sebesar 92,9%, 95,2% dan 95,6%. Waktu komputasi yang dibutuhkan dalam proses *running* CSVM *kernel linear* pada tiga provinsi tersebut adalah 34,029 detik, 2,67 detik dan 6,054 detik.
8. Hasil ketepatan klasifikasi menggunakan jenis data frekuensi *tweet netizen* memberikan nilai akurasi lebih tinggi dibandingkan dengan menggunakan data frekuensi relatif pada metode *kernel linear* dan RBF di Provinsi Jawa Timur, Jawa Tengah dan Jawa Barat.
9. Hasil analisis klaster secara global tidak dapat merefleksikkan keadaan politik di dunia real kecuali pada wilayah dengan karakteristik tertentu.
10. Hasil visualisasi *Bubble Charts* menunjukkan bahwa banyaknya *tweet netizen* tidak dapat merefleksikan keadaan politik di dunia real, tapi dapat direfleksikan oleh banyaknya *netizen* yang *tweet* di twitter.

5.2 Saran

1. Pada penelitian klasifikasi *tweet* ini sangat membutuhkan komputer dengan daya *Random Access Memory* (RAM) yang cukup besar agar data *tweet netizen* seluruhnya dapat dianalisis. Sehingga hasil analisis lebih akurat.
2. Untuk penelitian selanjutnya, penelitian serupa dapat dikembangkan dengan menggunakan API *Stream Premium* dan dapat dibuat program untuk otomatisasi klasifikasi. Sehingga hasil analisis sentimen dapat diakses secara real *time* dan lebih akurat. Selain itu, daftar kata pada *stopwords* dapat dilengkapi dengan daftar kata singkatan dan daftar kata slang dalam bahasa Indonesia.

(*Halaman ini sengaja dikosongkan*)

DAFTAR PUSTAKA

Arifin, Z. A., Mahendra, I. P., & Ciptaningtyas, H. T. (2009). *Enhanced Confix Stripping Stemmer And Ants Algorithm For Classifying News Document In Indonesian Language*. 149. 2085-1944.

BPS. (2018). *Sensus Penduduk 2010*. Diakses pada 28 Maret 2018, dari BPS: http://sp2010.bps.go.id/ index.php/ site/index

Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press.

Flora, M. (2018). *Suara Khofifah-Emil Tak Terkejar di Real Count KPUD Jatim*. Diakses pada 10 Juli 2018, dari liputan6: https://www.liputan6.com/pilkada/read/ 3574926/suara-khofifah-emil-tak-terkejar-di-real-count-kpud-jatim

Gu, Q., & Han, J. (2013). Clustered Support Vector Machines. *Journal of Machine Learning Research*, 307-315.

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques*. USA: Morgan Kaufmann Publishers.

Herbrich, R., & Graepel, T. (2010). *Natural Language Processing*. United States of America: Taylor and Francis Group, LLC

Hidayat, W. (2014). *Pengguna Internet Indonesia Nomor Enam Dunia*. Diakses pada 22 Februari 2018, dari kompas: http://tekno.kompas.com/ read/2014/11/24/ 07430087/ Pengguna. Internet.Indonesia. Nomor. Enam. Dunia.

Hardle, W. K., Prastyo, D. D., & Hafner, C. (2014). *Support Vector Machines with Evolutionary Model Selection for Default Prediction*. Dalam J. Racine, L. Su, & A. Ullah, *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics* (hal. 346-373). New York: OXFORD University Press.

Kumar, L., & Bhatia, P. K. (2013). Text Mining: Concepts, Proses and Applications. *Journal of Global Research in Computer Science* , 36-39.

Kusuma, E. F. (2015). *Bagaimana Peran Twitter Mempengaruhi Politik Indonesia*. Diakses pada 22 Februari 2018, dari detik:https://inet.detik.com/cyberlife/d-2943830/bagaimana-peran-twitter-mempengaruhi-politik-Indonesia.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Boca Raton: Morgan & Claypool Publisher.

Maulana, A. (2016). *Twitter Rahasiakan jumlah Pengguna di Indonesia.* Diakses pada 22 Februari 2018, dari CNN: https://www.cnnindonesia.com/teknologi/20160322085045-185-118939/twitter-rahasiakan-jumlah-pengguna-di-indonesia/.

Mardiastuti, A. (2018). *Ini Daftar 116 Cagub Cawagub di 17 Provinsi*. Diakses pada 22 Februari 2018, dari detik: https://news.detik.com/berita/d-3809876/ini-daftar-116-cagub-cawagub-di-17-provinsi.

Muzani, S. (2018). *Tahun Politik 2018*: *Kekuatan Partai dan Calon Presiden*. (R. Apinino, Pewawancara).

Nuansa, E. P., (2017). Analsis Sentimen Pengguna Twitter Terhadap Pemilihan Gubernur DKI Jakarta Dengan Metode Naïve Bayesian Classification Dan Support Vector Machine. Surabaya: Institut Teknologi Sepuluh Nopember.

Pozzi, F. A., Fersini, E., Messina, E., & Liu, B. (2017). *Sentiment Analysis in Social Networks*. United States: ELSEVIER.

Prasetyo, E. (2012) *Data Mining: Konsep dan Aplikasi menggunakan Matlab*, 1 ed. Yogyakarta: Andi Offset.

Purbaya, A. A. (2018). *Hasil Rekapitulasi KPU Jateng, Ganjar-Yasin Unggul*. Diakses pada 10 Juli 2018, dari detik: https://news.detik.com/jawatengah/4104351/hasil-rekapitulasi -kpu-jateng-ganjar-yasin-unggul.

Ramdhani, D. (2018). *Rapat Pleno KPU Jabar, Ridwan Kamil-Uu Ruzhanul Menangi Pilkada Jabar*. Diakses pada 10 Juli 2018, dari Kompas: https://regional.kompas.com/read/ 2018/07/08/16145081/rapat-pleno-kpu-jabar-ridwan-kamil-uu-ruzhanul-menangi-pilkada-jabar.

Sheth, A. (2013). Semantic Web Ontology and Knowledge Base Enabled Tools, Services, and Application. United States. of America: *Information Science Reference of IGI Global*

Smith, A. (2011). *Twitter and Social Networking in the 2010 Midterm Elections. Pew Internet and American Life* Diakses pada 22 Februari 2018, dari pewresearch: http://pewresearch.org/pubs/1871 /internetpolitics-facebook-twitter-2010-midterm-elections-campaign.

Statistika. (2018). *Number of Monthly Active Twitter Users*. Diakses pada 22 Februari 2018, dari statista: https://www.statista.com: https://www.statista.com/ statistics /282087/number-of-monthly-active-twitter-users/.

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)*, 421431.

Trivedi, S. K., Dey, S., Kumar, A., & Panda, T. K. (2017). *Handbook of Research on Advanced Data Mining Techniques and Applications for Business Intelegence*. United States of America: IGI Gobal.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sen*time*nt. *In Proceedings of ICWSM'2010*.

Weiss, S. M. (2010). *Text mining: Predictive Methods for Analyzing Unstructured Information*. New York: Springer.

Witten, I. H., Eibe, F., & Hall, M. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. United States: Morgan Kaufmann.

Xie, J., Wang, C., Zhang, Y., & Jiang, S. (2009). Clustering Supoprt Vector Machines for Unlabeled Data Classification. *International Conference on Test and Measurement* , 34-38.

LAMPIRAN

Lampiran 1. Ketepatan Klasifikasi Data Frekuensi *Tweet Netizen* Menggunakan CSVM *Kernel Linear*.

a. Provinsi Jawa Timur

C	Akurasi	G-Mean	AUC	*Time*
0,001	0,9989	0,9977	0,9977	15,02
0,01	1	1	1	8,283
0,1	1	1	1	7,308
1	1	1	1	7,157
10	1	1	1	7,189
100	1	1	1	7,272
1000	1	1	1	7,207

b. Provinsi Jawa Tengah

C	Akurasi	G-Mean	AUC	*Time*
0,001	0,9786	0,9707	0,9712	1,99
0,01	0,9786	0,9707	0,9712	1,879
0,1	0,9786	0,9707	0,9712	1,752
1	0,9786	0,9707	0,9712	1,731
10	0,9786	0,9707	0,9712	1,751
100	0,9786	0,9707	0,9712	1,822
1000	0,9786	0,9707	0,9712	1,807

Lampiran 1. Ketepatan Klasifikasi Data Frekuensi *Tweet Netizen* Menggunakan CSVM *Kernel Linear* (Lanjutan).

c. Provinsi Jawa Barat

C	**Akurasi**	**G-Mean**	**AUC**	***Time***
0,001	0,9870	0,8859	0,8929	2,535
0,01	0,9870	0,8859	0,8929	2,458
0,1	0,9870	0,8859	0,8929	2,487
1	0,9870	0,8859	0,8929	2,329
10	0,9870	0,8859	0,8929	2,413
100	0,9870	0,8859	0,8929	2,369
1000	0,9870	0,8859	0,8929	2,501

Lampiran 2. Ketepatan Klasifikasi Data Frekuensi Relatif *Tweet Netizen* Menggunakan CSVM *Kernel Linear*.

a. Provinsi Jawa Timur

C	Akurasi	G-Mean	AUC	*Time*
0,001	0,9293	0,8689	0,8754	62,815
0,01	0,9293	0,8689	0,8754	57,856
0,1	0,9294	0,8690	0,8755	57,745
1	0,9349	0,8877	0,8918	56,796
10	0,9410	0,9046	0,9071	48,072
100	0,9456	0,9204	0,9216	31,226
1000	0,9412	0,9235	0,9241	19,66

b. Provinsi Jawa Tengah

C	Akurasi	G-Mean	AUC	*Time*
0,001	0,9310	0,9043	0,9098	4,286
0,01	0,9310	0,9043	0,9098	4,304
0,1	0,9452	0,9253	0,9287	4,154
1	0,9714	0,9606	0,9619	3,39
10	0,9667	0,9548	0,9562	3,327
100	0,9619	0,9481	0,9500	3,129
1000	0,9619	0,9481	0,9500	3,037

Lampiran 2. Ketepatan Klasifikasi Data Frekuensi Relatif *Tweet Netizen* Menggunakan CSVM *Kernel Linear* (Lanjutan).

c. Provinsi Jawa Barat

C	**Akurasi**	**G-Mean**	**AUC**	***Time***
0,001	0,9293	0,6661	0,7115	5,593
0,01	0,9293	0,6661	0,7115	5,547
0,1	0,9293	0,6661	0,7115	5,46
1	0,9303	0,6668	0,7122	4,609
10	0,9234	0,6624	0,7082	3,556
100	0,9234	0,6630	0,7086	3,278
1000	0,9224	0,6623	0,7079	3,366

Lampiran 3. Ketepatan Klasifikasi Data Frekuensi *Tweet Netizen* Provinsi Jawa Timur Menggunakan CSVM *Kernel* RBF.

C	Gamma	Akurasi	G-Mean	AUC	*Time*
0,01	0,01	0,8145	0,4235	0,5899	42,977
0,01	0,1	0,9358	0,8544	0,8644	56,553
0,01	1	0,9710	0,9756	0,9756	59,914
0,01	10	0,9284	0,9523	0,9534	55,46
0,01	100	0,7121	0,7925	0,8140	45,41
0,1	0,01	0,9990	0,9988	0,9988	22,657
0,1	0,1	0,9954	0,9965	0,9965	35,005
0,1	1	0,9740	0,9824	0,9825	50,871
0,1	10	0,9284	0,9523	0,9534	46,822
0,1	100	0,7121	0,7925	0,8140	39,465
1	0,01	0,9998	0,9997	0,9997	13,447
1	0,1	0,9962	0,9974	0,9974	25,7
1	1	0,9743	0,9828	0,9830	35,968
1	10	0,9285	0,9523	0,9534	33,332
1	100	0,7117	0,7921	0,8138	28,809
10	0,01	0,9998	0,9997	0,9997	13,211
10	0,1	0,9962	0,9974	0,9974	25,12
10	1	0,9743	0,9828	0,9830	35,872
10	10	0,9285	0,9523	0,9534	33,038
10	100	0,7117	0,7921	0,8138	28,663
100	0,01	0,9998	0,9997	0,9997	12,795
100	0,1	0,9962	0,9974	0,9974	25,055
100	1	0,9743	0,9828	0,9830	35,723
100	10	0,9285	0,9523	0,9534	33,068
100	100	0,7117	0,7921	0,8138	28,777

Lampiran 4. Ketepatan Klasifikasi Data Frekuensi *Tweet Netizen* Provinsi Jawa Tengah Menggunakan CSVM *Kernel* RBF.

C	Gamma	Akurasi	G-Mean	AUC	*Time*
0,01	0,01	0,9381	0,9214	0,9255	3,123
0,01	0,1	0,8357	0,7854	0,8105	2,848
0,01	1	0,5571	0,0755	0,5146	2,872
0,01	10	0,5381	0,0000	0,5000	2,762
0,01	100	0,5381	0,0000	0,5000	2,785
0,1	0,01	0,9381	0,9214	0,9255	2,854
0,1	0,1	0,8357	0,7854	0,8105	2,684
0,1	1	0,5571	0,0755	0,5146	3,087
0,1	10	0,5381	0,0000	0,5000	2,89
0,1	100	0,5381	0,0000	0,5000	2,682
1	0,01	0,9381	0,9214	0,9255	2,754
1	0,1	0,8357	0,7854	0,8105	3,048
1	1	0,5571	0,0755	0,5146	2,727
1	10	0,5381	0,0000	0,5000	2,567
1	100	0,5381	0,0000	0,5000	2,65
10	0,01	0,9381	0,9214	0,9255	2,729
10	0,1	0,8357	0,7854	0,8105	2,803
10	1	0,5571	0,0755	0,5146	2,837
10	10	0,5381	0,0000	0,5000	2,803
10	100	0,5381	0,0000	0,5000	2,796
100	0,01	0,9381	0,9214	0,9255	2,809
100	0,1	0,8357	0,7854	0,8105	2,76
100	1	0,5571	0,0755	0,5146	2,804
100	10	0,5381	0,0000	0,5000	3,067
100	100	0,5381	0,0000	0,5000	2,767

Lampiran 5. Ketepatan Klasifikasi Data Frekuensi *Tweet Netizen* Provinsi Jawa Barat Menggunakan CSVM *Kernel* RBF.

C	Gamma	Akurasi	G-Mean	AUC	*Time*
0,01	0,01	0,9731	0,8124	0,8301	6,6890
0,01	0,1	0,9323	0,7469	0,7753	6,7300
0,01	1	0,4342	0,5697	0,6175	6,6210
0,01	10	0,0279	0,4351	0,5018	6,4290
0,01	100	0,0209	0,4330	0,5000	6,3870
0,1	0,01	0,9731	0,8124	0,8301	6,4050
0,1	0,1	0,9323	0,7469	0,7753	6,4330
0,1	1	0,4382	0,5712	0,6187	6,3070
0,1	10	0,0279	0,4351	0,5018	6,2440
0,1	100	0,0209	0,4330	0,5000	6,3450
1	0,01	0,9731	0,8124	0,8301	6,4830
1	0,1	0,9323	0,7469	0,7753	6,5270
1	1	0,4382	0,5712	0,6187	6,3390
1	10	0,0279	0,4351	0,5018	6,3470
1	100	0,0209	0,4330	0,5000	6,4210
10	0,01	0,9731	0,8124	0,8301	6,5290
10	0,1	0,9323	0,7469	0,7753	6,3120
10	1	0,4382	0,5712	0,6187	6,4390
10	10	0,0279	0,4351	0,5018	6,3510
10	100	0,0209	0,4330	0,5000	6,2990
100	0,01	0,9731	0,8124	0,8301	6,4100
100	0,1	0,9323	0,7469	0,7753	6,5560
100	1	0,4382	0,5712	0,6187	6,3040
100	10	0,0279	0,4351	0,5018	6,3360
100	100	0,0209	0,4330	0,5000	6,5240

Lampiran 6. Ketepatan Klasifikasi Data Frekuensi Relatif *Tweet Netizen* Provinsi Jawa Timur Menggunakan CSVM *Kernel* RBF.

C	Gamma	Akurasi	G-Mean	AUC	*Time*
0,01	0,01	0,7489	0,0000	0,5000	50,0210
0,01	0,1	0,7489	0,0000	0,5000	50,5850
0,01	1	0,7489	0,0000	0,5000	52,6060
0,01	10	0,7492	0,0188	0,5006	54,7890
0,01	100	0,7559	0,1827	0,5159	58,7310
0,1	0,01	0,7489	0,0000	0,5000	43,1830
0,1	0,1	0,7489	0,0000	0,5000	43,4880
0,1	1	0,7491	0,0110	0,5003	45,8730
0,1	10	0,7526	0,1233	0,5076	48,7760
0,1	100	0,8009	0,4690	0,6083	53,0370
1	0,01	0,7489	0,0000	0,5000	41,0140
1	0,1	0,7491	0,0110	0,5003	44,2780
1	1	0,7533	0,1334	0,5089	47,4210
1	10	0,8060	0,4871	0,6174	45,7100
1	100	0,8983	0,7889	0,8091	43,3000
10	0,01	0,7491	0,0110	0,5003	46,8870
10	0,1	0,7534	0,1355	0,5092	47,6870
10	1	0,8060	0,4867	0,6173	44,0380
10	10	0,8982	0,7936	0,8122	37,1510
10	100	0,9239	0,8693	0,8749	35,8760
100	0,01	0,7536	0,1376	0,5096	62,9700
100	0,1	0,8060	0,4868	0,6174	51,6160
100	1	0,8972	0,7936	0,8120	39,0060
100	10	0,9217	0,8690	0,8743	30,0600
100	100	0,9296	0,8911	0,8940	34,0290

Lampiran 7. Ketepatan Klasifikasi Data Frekuensi Relatif *Tweet Netizen* Provinsi Jawa Tengah Menggunakan CSVM *Kernel* RBF.

C	Gamma	Akurasi	G-Mean	AUC	*Time*
0,01	0,01	0,6310	0,0000	0,5000	3,282
0,01	0,1	0,6310	0,0000	0,5000	3,179
0,01	1	0,6357	0,0539	0,5073	3,106
0,01	10	0,6762	0,3461	0,5640	3,075
0,01	100	0,6810	0,5096	0,6288	2,966
0,1	0,01	0,6310	0,0000	0,5000	2,605
0,1	0,1	0,6310	0,0000	0,5000	2,732
0,1	1	0,6452	0,1490	0,5230	2,782
0,1	10	0,6762	0,3461	0,5640	3,065
0,1	100	0,6810	0,5096	0,6288	3,11
1	0,01	0,6310	0,0000	0,5000	2,561
1	0,1	0,6786	0,3556	0,5684	2,604
1	1	0,9119	0,8724	0,8817	2,693
1	10	0,7524	0,5682	0,6655	2,823
1	100	0,6810	0,5096	0,6288	2,947
10	0,01	0,6881	0,3828	0,5799	2,537
10	0,1	0,9214	0,8853	0,8927	2,375
10	1	0,9524	0,9353	0,9379	2,67
10	10	0,7714	0,6151	0,6944	2,848
10	100	0,6976	0,6038	0,6847	3,032
100	0,01	0,9214	0,8853	0,8927	2,562
100	0,1	0,9476	0,9313	0,9337	2,491
100	1	0,9500	0,9336	0,9362	2,839
100	10	0,8048	0,6968	0,7464	3,132
100	100	0,6262	0,6237	0,6911	3,191

Lampiran 8. Ketepatan Klasifikasi Data Frekuensi Relatif *Tweet Netizen* Provinsi Jawa Barat Menggunakan CSVM *Kernel* RBF.

C	Gamma	Akurasi	G-Mean	AUC	*Time*
0,01	0,01	0,5627	0,4330	0,5000	7,787
0,01	0,1	0,5627	0,4330	0,5000	7,764
0,01	1	0,5647	0,4342	0,5011	7,755
0,01	10	0,7377	0,5547	0,6066	7,769
0,01	100	0,8281	0,6258	0,6688	7,872
0,1	0,01	0,5627	0,4330	0,5000	6,841
0,1	0,1	0,5627	0,4330	0,5000	6,955
0,1	1	0,5716	0,4388	0,5050	7,14
0,1	10	0,7943	0,5946	0,6419	7,683
0,1	100	0,8281	0,6258	0,6688	7,786
1	0,01	0,5627	0,4330	0,5000	6,913
1	0,1	0,5667	0,4354	0,5021	6,954
1	1	0,9394	0,7162	0,7486	6,613
1	10	0,9474	0,7434	0,7709	6,896
1	100	0,9086	0,6929	0,7279	7,816
10	0,01	0,5667	0,4354	0,5021	7,168
10	0,1	0,9424	0,7234	0,7546	6,718
10	1	0,9543	0,7598	0,7841	6,185
10	10	0,9444	0,7394	0,7676	6,925
10	100	0,9116	0,6971	0,7315	7,795
100	0,01	0,9434	0,7282	0,7584	6,764
100	0,1	0,9563	0,7657	0,7889	6,054
100	1	0,9464	0,7445	0,7714	6,165
100	10	0,9434	0,7387	0,7670	6,858
100	100	0,9116	0,6971	0,7315	7,835

Lampiran 9. *Syntax Crawling* Data Menggunakan R-Studio.

```
#Twitter ID API
api_key = "your api_key"
api_secret =your api_secret"
access_token = "your acces_token"
access_token_secret = "youracces_toke_secret"

#Autorize Log in Twitter ID
setup_twitter_oauth(api_key, api_secret, access_token,
                    access_token_secret)

#Searching Tweets
tvone <- searchTwitter("@ganjarpranowo", lang="id", n=100000,
                    resultType="recent")
write,csv(twListToDF(tvone), file="ganjarpranowo21Juni,csv")
```

Lampiran 10. *Syntax Import* Data dan Praproses Data Menggunakan Python 2,7.

```
import pandas as pd
import string
import nltk
import re
import sys
import os
import glob
import numpy as numpy
from Sastrawi,Stemmer,StemmerFactory import
    StemmerFactory
#Import Data
path = r'D:/Kuliah/S1 Statistika ITS/Semester VIII/Data
    Crawling Terbaru/Wilayah/Jawa Timur
globbed_files = glob,glob(os,path,join(path, "*,csv"))
data = []
   frame = pd,read_csv(csv)
   frame['filename'] = os,path,basename(csv)
   data,append(frame)
bigframe = pd,concat(data, ignore_index=True)
bigframe,shape
bigframe,drop(bigframe,columns,difference(['text','screenName
    ','filename']), 1, inplace=True)
bigframe=bigframe,drop_duplicates()
bigframe=bigframe,dropna()
bigframe,shape
df=bigframe
df,shape
```

Lampiran 10. *Syntax Import* Data dan Praproses Data Menggunakan Python 2,7 (Lanjutan).

```
df['Cagub 1'] = numpy,where(df['filename'],str[:1]=="r", 1, 0)
df['Wagub 1'] = numpy,where(df['filename'],str[:1]=="u", 1, 0)
df['Cagub 2'] = numpy,where(df['filename'],str[:1]=="t", 1, 0)
df['Wagub 2'] = numpy,where(df['filename'],str[:1]=="A", 1, 0)
df['Cagub 3'] = numpy,where(df['filename'],str[:1]=="M", 1, 0)
df['Wagub 3'] = numpy,where(df['filename'],str[:1]=="s", 1, 0)
df['Cagub 4'] = numpy,where(df['filename'],str[:4]=="Dedd", 1,
                0)
df['Wagub 4'] = numpy,where(df['filename'],str[:4]=="Dedi", 1,
                0)
df,drop(df,columns[[2]], axis=1, inplace=True)
df['Cagub 1'] = numpy,where(df['filename'],str[:1]=="K", 1, 0)
df['Wagub 1'] = numpy,where(df['filename'],str[:1]=="E", 1, 0)
df['Cagub 2'] = numpy,where(df['filename'],str[:1]=="g", 1, 0)
df['Wagub 2'] = numpy,where(df['filename'],str[:1]=="P", 1, 0)
df,drop(df,columns[[2]], axis=1, inplace=True)
df,head()
df,shape
text_train = df['text'] #ambil kolom text
df_train = text_train

# Menghapus Link
trainnolink = []
for line in df_train:
   result = re,sub(r"http\S+", "", line)
   trainnolink,append(result)
```

Lampiran 10. *Syntax Import* Data dan Praproses Data Menggunakan Python 2,7 (Lanjutan).

```
#Menghapus Re*tweet*
trainnort = []
for line in trainnolink:
    result = re,sub(r"RT", "", line)
    trainnort,append(result)

#Menghapus Username
trainnousername = []
for line in trainnort:
    result = re,sub(r"@\S+", "", line)
    trainnousername,append(result)

#Case Folding
train_lower = []
for line in trainnousername:
    a = line,lower()
    train_lower,append(a)

#Stemming
factory = StemmerFactory()
stemmer = factory,create_stemmer()
train_stemmed = map(lambda x: stemmer,stem(x), train_lower)
train_no_punc = map(lambda x: x,lower(),translate(None, string,punctuation), train_stemmed)
```

Lampiran 10. *Syntax Import* Data dan Praproses Data Menggunakan Python 2,7 (Lanjutan).

```
#Filtering
stopword = open("D:/stopword,txt","r"),read()
trainfinal = []
for line in train_no_punc:
   word_token = nltk,word_tokenize(line) # get word token forevery line (split line into each separate words)
   word_token = [word for word in word_token if not word in stopword and not word[0],isdigit()] # remove indonesian stop words and number
   trainfinal,append(" ",join(word_token))
   testfinal = []

#Menghitung Frekunsi per Kata
df1 = pd,DataFrame({'text':trainfinal})
df1,reset_index(drop=True, inplace=True)
df,reset_index(drop=True, inplace=True)
df2 = pd,concat([df1,df,iloc[:,1:6]], axis=1)
df2,head()
df2,shape
df3=df2,text,str,split(' ')
df3 = pd,DataFrame(df3,values,tolist())
df3,head()
df4=pd,DataFrame()
for i in range(df3,shape[1]):
   df4= pd,concat([df4,df3[i]])
df4=df4,rename(index=str, columns={0: "Karakteristik"})
df4=df4,dropna(how='all')
df5=df4['Karakteristik'],value_counts()
df5=df4['Karakteristik'],value_counts()
df6 =
df5,rename_axis('Karakteristik'),reset_index(name='Frekuensi')
```

Lampiran 10. *Syntax Import* Data dan Praproses Data Menggunakan Python 2,7 (Lanjutan).

```
df6 =
df5,rename_axis('Karakteristik'),reset_index(name='Frekuensi')
df6,tail()
df7 = df6,drop(df6[df6,Frekuensi < 11],index)
df7,drop(df7,columns[[1]], axis=1, inplace=True)
df7=df7,dropna()
df7,head()
df7,to_csv('D:/jawatimur_karakteristik,csv',
sep=',',index=False)
df7,shape

# Menghitung Frekeuensi Relatif Kata
df100=df2[["Cagub 1","Wagub 1","Cagub 2", "Wagub
2"]]/len(df2)
dfb=pd,concat([df2["screenName"],df100], axis=1)
dfb,head()
dfa=pd,DataFrame(df2,iloc[:,1:6])
dfa,head()
dfc=dfa,groupby('screenName'),sum()
dfcc=dfb,groupby('screenName'),sum()
dfc,head()
dfcc,head()
def term1(row):
   return row['Cagub 1'] + row['Wagub 1']
dfc['Calon 1'] = dfc,apply(term1, axis=1)

def term2(row):
   return row['Cagub 2'] + row['Wagub 2']
dfc['Calon 2'] = dfc,apply(term2, axis=1)
dfc['label'] = numpy,where(dfc['Calon 1']>dfc['Calon 2'], 1, 2)
```

Lampiran 10. *Syntax Import* Data dan Praproses Data Menggunakan Python 2,7 (Lanjutan).

```
dfd = dfc[["label","Calon 1","Calon 2","Cagub 1","Wagub 1","Cagub 2","Wagub 2"]]
dfd,head()
dfdd=pd,concat([dfd,iloc[:,0:3],dfcc], axis=1)
dfdd,head()
df2,drop(df2,iloc[:,1:6], axis=1, inplace=True)
df2,head()
len(dfd)
len(df2)
s1=dict()
for i in range(df7,shape[0]):
   s1[i]=set(df7,iloc[i,0],split())
s2=dict()
for i in range(df2,shape[0]):
   s2[i]=set(df2,iloc[i,0],split())
fil=dict()
for i in range(df7,shape[0]):
   fil[i] = []
   for j in range (df2,shape[0]):
      fil[i],append(len(s2[j],intersection(s1[i])))
strukdat=dict()
for i in range(df7,shape[0]):
   strukdat[i]=pd,DataFrame()
   strukdat[i] = pd,DataFrame(fil[i])
strukdat[2],sum()
df8=pd,DataFrame()
for i in range(len(df7)):
   df8=pd,concat([df8,(strukdat[i])],axis=1)
df8,columns=['Kata '+str(i) for i in range (1,len(df7)+1)]
```

Lampiran 10. *Syntax Import* Data dan Praproses Data Menggunakan Python 2,7 (Lanjutan).

```
df9=pd,DataFrame()
for i in range(len(df7)):
   df9=pd,concat([df9,(strukdat[i])/(strukdat[i],sum())],axis=1)
## Dibagi dengan frekeunsi kata muncul dalam tweet netizen
df9,columns=['Kata '+str(i) for i in range (1,len(df7)+1)]
df10=pd,concat([df["screenName"],df8], axis=1)
df11=pd,concat([df["screenName"],df9], axis=1)
df12=df10,groupby('screenName'),sum()
df13=df11,groupby('screenName'),sum()
df14=pd,concat([dfd,df12], axis=1)
df15=pd,concat([dfdd,df13], axis=1)
df14,shape
df15,shape
df16=df14,loc[:, (df14 != 0),any(axis=0)]
df17=df15,loc[:, (df15 != 0),any(axis=0)]
df16,shape
df17,shape
df16,head()
df17,head()
df16,to_csv('Data Frekuensi Jawatimur,csv', sep=";",decimal=",")
df17,to_csv('Data FrekRelatif Jawatimur,csv',sep=";",decimal=",")
df17,drop(df17,iloc[:,1:3], axis=1, inplace=True)
df17,reset_index(inplace=True) # Resets the index, makes factor a column
df17,drop("screenName",axis=1,inplace=True)
```

Lampiran 10. *Syntax Import* Data dan Praproses Data Menggunakan Python 2,7 (Lanjutan).

```
df18=df15,loc[:, (df15 != 0),any(axis=0)]
df18,reset_index(inplace=True) # Resets the index, makes factor a column
df18,drop(df18,columns[[0,2,3]], axis=1, inplace=True)
df18,head()
df18,to_csv('data kmeans,csv', index=False)
```

Lampiran 11. *Syntax* Klasifikasi Menggunakan Metode CSVM.

```
library(caret)
library(SwarmSVM)
training <- read,csv(file="C:/Users/M As'adur Rofiq/data kameans,csv", header=TRUE, sep=",")
folds <- createFolds(factor(training$label), k = 10, list = TRUE,returnTrain=TRUE)
str(folds)
data <-training[folds$Fold03,]
nrow(data)
trainRowNumbers <- createDataPartition(data$label, p=0,8, list=FALSE)
trainData <- data[trainRowNumbers,]
testData <- data[-trainRowNumbers,]
svmguide1,t = as,matrix(testData)
svmguide1 = as,matrix(trainData)
svmguide1[,1:5]
b=0,01
for(i in 1:5){
  for(j in 1:5){
    if(j==1){
      a=0,01
    }
    else if(j==2){
      a=0,1
    }
    else if(j==3){
      a=1
    }
    else if(j==4){
      a=10
    }
```

Lampiran 11. *Syntax* Klasifikasi Menggunakan Metode CSVM (Lanjutan).

```
}
   else {
    a=100
   }

   dcsvm,model = dcSVM(x = svmguide1[,-1], y = svmguide1[,1],verbose = FALSE,
                  k = 2, max,levels = 4, seed = 512, cost = b, gamma = a,
                  kernel = 1,early = 0, m = 800,
                  valid,x = svmguide1,t[,-1], valid,y = svmguide1,t[,1])
   d=dcsvm,model$valid,score ## Akurasi
   c=dcsvm,model$Time$total,Time ## Time process total
   preds = dcsvm,model$valid,pred
   e=table(preds,svmguide1,t[,1]) ##confusion matrix
   f=e[1,1]/(e[1,1]+e[1,2]) ##sensityfity
   g=e[2,2]/(e[2,2]+e[2,1]) ##specificity
   gmean=sqrt(f*g) ##gmean
   AUC=(f*g)/2 ##AUC
   print=cat("Akurasi:",d,"C:",b,"Gamma:",a,"G-Mean:",gmean,"AUC:",AUC,"Time:",c,"\n")

  }
  b=b*10
}
```

Lampiran 12. *Syntax Word cloud*.

```
#Install and load the required packages

install,packages("tm")  # for text mining
install,packages("SnowballC") # for text stemming
install,packages("wordcloud") # word-cloud generator
install,packages("RColorBrewer") # color palettes
library("tm")
library("SnowballC")
library("wordcloud")
library("RColorBrewer")
library("wordcloud2")
#Text mining
text <- readLines("D:/Data Preprocessing/Karakteristik Klaster/Jawa Timur/Klaster1,txt")
docs <- Corpus(VectorSource(text))

# Text Transformation

toSpace <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))
docs <- tm_map(docs, toSpace, "/")
docs <- tm_map(docs, toSpace, "@")
docs <- tm_map(docs, toSpace, "\\|")

# Cleaning the text

# Convert the text to lower case
docs <- tm_map(docs, content_transformer(tolower))
# Remove numbers
docs <- tm_map(docs, removeNumbers)
# Remove english common stopwords
```

Lampiran 12. *Syntax Word cloud* (Lanjutan).

```
# Remove english common stopwords
docs <- tm_map(docs, removeWords, stopwords("english"))
# Remove your own stop word
# specify your stopwords as a character vector
docs <- tm_map(docs, removeWords, c("null","grobogan","ekotren"))
# Remove punctuations
docs <- tm_map(docs, removePunctuation)
# Eliminate extra white spaces
docs <- tm_map(docs, stripWhitespace)
# Text stemming
docs <- tm_map(docs, stemDocument)


# Build a term-document matrix

dtm <- TermDocumentMatrix(docs)
m <- as,matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data,frame(word = names(v),freq=v)
head(d, 20)

#Generate the Word cloud

set,seed(1234)
wordcloud(words = d$word, freq = d$freq, min,freq = 1,
        max,words=100000, random,order=FALSE, rot,per=0,35,
        colors=brewer,pal(8, "Dark2"))
```

Lampiran 13. Surat Pernyataan Data.

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini, mahasiswa Departemen Statistika FMKSD ITS:

Nama : M. Asadur Rofiq

NRP : 06211440000097

menyatakan bahwa data yang digunakan dalam Tugas Akhir/ Thesis ini merupakan data sekunder yang diambil dari ~~penelitian /buku/ Tugas Akhir/ Thesis/~~ publikasi lainnya yaitu:

Sumber : Situs Resmi *infopemilu.kpu.go.id*

Keterangan : Hasil Rekapitulasi Suara Pemilihan Kepala Daerah Provinsi Jawa Timur, Jawa Tengah, dan Jawa Barat

Surat Pernyataan ini dibuat dengan sebenarnya. Apabila terdapat pemalsuan data maka saya siap menerima sanksi sesuai aturan yang berlaku.

Surabaya, 24 Juli 2018

Mengetahui,
Co- Pembimbing Tugas Akhir

Dr.rer.pol. Dedy Dwi Prastyo, M.Si
NIP. 19831204 200812 1 002

M. Asadur Rofiq
NRP. 06211440000097

Pembimbing Tugas Akhir

Imam Safawi Ahmad, S.Si., M.Si.
NIP. 19810224 201404 1 001

Lampiran 13. Surat Pernyataan Data (Lanjutan).

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini, mahasiswa Departemen Statistika FMKSD ITS:

Nama : M. Asadur Rofiq

NRP : 06211440000097

menyatakan bahwa data yang digunakan dalam Tugas Akhir/ Thesis ini merupakan data primer yang diambil dari:

Sumber : Halaman Twitter

Keterangan : *tweet* pengguna twitter di Indonesia terhadap akun twitter Calon Kepala Daerah dan Wakil Provinsi Jawa Timur, Jawa Tengah, dan Jawa Barat

Surat Pernyataan ini dibuat dengan sebenarnya. Apabila terdapat pemalsuan data maka saya siap menerima sanksi sesuai aturan yang berlaku.

Surabaya, 24 Juli 2018

Mengetahui,
Co- Pembimbing Tugas Akhir

Dr.rer.pol. Dedy Dwi Prastyo, M.Si
NIP. 19831204 200812 1 002

M. Asadur Rofiq
NRP. 06211440000097

Pembimbing Tugas Akhir

Imam Safawi Ahmad, S.Si., M.Si.
NIP. 19810224 201404 1 001

(*Halaman ini sengaja dikosongkan*)

BIODATA PENULIS

Penulis lahir di Lumajang, 29 Januari 1995 dengan nama lengkap Mochammad As'adur Rofiq dan biasa dipanggil dengan nama Aat Jobs. Penulis merupakan anak pertama dari pasangan Bapak Komari dan Ibu Sanah Nurchasanah. Pendidikan formal penulis diawali di TK Islami Jatisari 1 - Lumajang (tahun 2000-2001), SD Negeri Jatisari 1 - Lumajang (tahun 2001-2007), SMP Negeri 1 Tempeh - Lumajang (tahun 2007-2010), MA Nurul Jadid Paiton - Probolinggo (tahun 2010-2013), hingga diterima di S1 Statistika Institut Teknologi Sepuluh Nopember (ITS) Surabaya pada tahun 2014. Selama menempuh pendidikan di jenjang kuliah, penulis tergabung dalam UKM Penalaran ITS selama tiga periode dan lembaga dakwah jurusan FORSIS-ITS sebagai staff Kaderisasi. Sedangkan, diluar kampus penulis juga pernah aktif dalam organisasi Lembaga Amil Zakat dan Infaq sebagai pelaksana Bimbingan Belajar SBMTN selama empat periode dan Pergerakan Mahasiswa Islam Indonesia (PMII 1011) sebagai ketua Penelitian dan Pengembangan selama empat periode. Selain itu penulis juga aktif dalam kegiatan sosial dalam memberikan motivasi dan bimbingan belajar suskes lanjut ke perguruan tinggi di beberapa pesantren di Jawa Timur. Jika ingin berdiskusi lebih lanjut mengenai Tugas Akhir penulis, dapat menghubungi penulis melalui email berikut: mochammadasadurrofiq@gmail.com.