

Penerapan Sentimen Analisis Acara Televisi Pada Twitter Menggunakan Support Vector Machine dan Algoritma Genetika sebagai Metode Seleksi Fitur

I Made Budi Surya Darma¹, Rizal Setya Perdana², Indriati³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹budisuryadarma0407@gmail.com, ²rizalespe @ub.ac.id, ³indriati.tif@ub.ac.id

Abstrak

Rating merupakan salah satu metode pendekatan yang dapat digunakan untuk mengetahui tingkat kepuasan penonton terhadap suatu program acara TV. Di Indonesia, perhitungan *rating* dilakukan dengan menggunakan jasa AGB Nielsen yakni berdasarkan perilaku menonton dari masyarakat di 10 kota besar di Indonesia. Karena itu, diperlukan metode pengamatan lainnya untuk mendapatkan perilaku menonton dari keseluruhan masyarakat di Indonesia. Twitter, dapat dimanfaatkan untuk mengamati perilaku menonton masyarakat Indonesia. Melalui tweet yang dipublikasikan, dapat dilakukan proses penggalian informasi dengan menggunakan teknik klasifikasi untuk mendapatkan opini dari masyarakat mengenai suatu acara. Salah satu teknik klasifikasi yang dapat diterapkan pada kategorisasi teks adalah *Support Vector Machine* (SVM), dimana SVM cocok untuk mengatasi data yang berdimensi banyak. Dengan mengoptimalkan fitur yang digunakan pada SVM, dapat mengurangi beban kerja namun, tetap memberikan hasil yang optimal. Salah satu metode seleksi fitur yang dapat diterapkan pada SVM adalah algoritma genetika (GA). Sistem menghitung *rating* berdasarkan tweet bersentimen positif dan negatif yang tertuju ke suatu acara dibagi dengan populasi dari tweet yang digunakan. Pengujian perbandingan *rating* yang dihasilkan sistem dengan *rating* AGB Nielsen menunjukkan rata-rata nilai *error* 0.562. Sedangkan pada pengujian akurasi sebelum dan sesudah penerapan metode seleksi fitur menunjukkan hasil dengan rata-rata nilai *error* 0.62%.

Kata kunci: sentimen analisis, *support vector machine*, algoritma genetika, seleksi fitur

Abstract

Rating is one approach method that can be used to find out about audience satisfaction of a TV show. In Indonesia, rating was calculated by using AGB Nielsen services. However, rating that AGB Nielsen produced was based on the people watching behavior in 10 major cities in Indonesia. Therefore, rating in Indonesia requires another method to get the watching behavior of the whole people in Indonesia. Twitter, can be used to get Indonesia people watching behavior. Through the published tweets, it can be applied the process of extracting information by using classification techniques to get the opinions. One of the classification techniques that can be applied to text categorization is the Support Vector Machine (SVM) it's suitable for multiple dimension data. By optimizing the features that will be used, it can provide optimal results with less features used. One of the feature selection methods that can be applied to SVM is the genetic algorithm (GA). System calculates the rating, based on positive and negative sentiments about the TV show and divided by the population of the tweet used. The rating comparison test that produced by AGB Nielsen and system shows an average error value of 0.562. In testing the accuracy before and after the feature selection method is applied, showed results with average error value 0.62%.

Keywords: *opinion mining, support vector machine, genetic algorithm, feature selection*

1. PENDAHULUAN

Rating merupakan salah satu pendekatan yang dapat dilakukan untuk mengetahui kepuasan penonton terhadap suatu program

acara televisi yang baru saja ditonton. Melalui *rating*, stasiun televisi dapat melakukan evaluasi terhadap suatu program acara. Program acara dengan *rating* yang rendah dapat diartikan bahwa program acara tersebut

mengalami penurunan jumlah penonton sehingga stasiun TV dapat mengambil tindakan dengan mengganti program acara tersebut dengan program acara yang baru. Di Indonesia, beberapa stasiun TV nasional melakukan perhitungan rating suatu acara dengan menggunakan jasa dari AGB Nielsen Media Research. Terdapat 10 stasiun TV nasional yang menggunakan jasa AGB Nielsen Media Research yakni RCTI, SCTV, Trans TV, Indosiar, MNC TV, Trans TV, Global TV, ANTV, TV ONE, dan Metro TV (Supardi, 2012). Kekurangan dari rating yang dihasilkan oleh AGB Nielsen ialah, rating yang dihasilkan merupakan perilaku menonton dari masyarakat yang ada pada 10 kota besar yang ada di Indonesia, sehingga belum mampu menggambarkan perilaku menonton dari masyarakat di luar 10 kota besar yang dimaksudkan dan, lebih lanjut lagi, menggambarkan perilaku menonton keseluruhan masyarakat Indonesia. Oleh karena itu, Pe-rating-an di Indonesia memerlukan metode pengamatan lainnya yang dapat dijadikan sarana untuk menggambarkan perilaku menonton dari keseluruhan masyarakat Indonesia. Salah satu media yang dapat dimanfaatkan adalah Twitter.

Twitter, merupakan salah satu situs media sosial dan *microblogging* terpopuler saat ini, khususnya dikalangan remaja, dimana user dapat saling bertukar informasi dengan menulis tweet, dengan penggunaan karakter yang dibatasi sebanyak 140 karakter (Fuvenanda, 2010). Menurut Capece, jaringan komunikasi dan penyebaran informasi di Twitter tergolong dinamis dalam memberikan pengaruh, sehingga informasi yang ada dapat berpengaruh hingga aspek psikologi, sosial, marketing, bisnis, hingga politik (Capece, 2009). Berdasarkan statistik yang dirilis oleh MIT Technology-Review pada tahun 2013, twitter menempati posisi keempat dengan nilai popularitas 22% sebagai salah satu media sosial paling berpengaruh berdasarkan jumlah pemakainya. Menurut MIT Technology-Review, dengan data statistik yang dirilis, pada tahun 2013 Indonesia berada pada posisi ketiga sebagai penyumbang tweet terbanyak dengan jumlah 1 milyar tweet, dimana posisi satu adalah Amerika Serikat (3,7 milyar) dan Jepang (1,8 milyar) pada posisi kedua (MIT, 2013).

Dengan banyaknya jumlah informasi yang disebar dengan menggunakan Twitter

sebagai sarannya, tidak sedikit dari informasi tersebut mengandung opini masyarakat mengenai suatu objek. Objek tersebut dapat berupa kejadian yang terjadi disekitar masyarakat, politik, atau bahkan suatu produk, salah satunya adalah acara TV. Banyak pengguna aktif Twitter yang berbagi opini, kritik atau saran melalui Twitter sesuai mereka menonton suatu program acara yang disiarkan. Melalui Twitter, pengguna (masyarakat) dapat melakukan diskusi dengan pengguna lainnya mengenai kepuasan, kekecewaan ataupun harapan terhadap acara TV yang baru saja ditonton sehingga dari melihat tweet yang dipublikasi oleh masyarakat, kita bisa mendapatkan gambaran mengenai opini masyarakat melalui proses penggalian informasi lebih lanjut dari tiap tweet yang dibuat oleh pengguna untuk mempelajari opini yang dituliskan. Proses analisis yang kerap digunakan untuk melihat opini masyarakat melalui tweet ialah sentimen analisis.

Sentimen Analisis (SA) atau *Opinion Mining* (OM) merupakan salah satu bidang penelitian dalam *text mining*. SA merupakan proses untuk melakukan identifikasi sentimen yang muncul pada suatu teks dengan mengolah data tekstual untuk memahami opini yang terkandung dalam suatu sentimen. SA cenderung dilakukan untuk melihat opini dari suatu sentimen terhadap sebuah objek dengan mengidentifikasi apakah sentimen tersebut mengandung opini positif atau negatif.

Dengan menggunakan sentimen analisis dapat dilakukan proses penggalian informasi untuk menganalisis opini yang terdapat pada tweet yang dituliskan oleh sekumpulan user terhadap suatu program acara TV. Dari hasil analisis opini tersebut dapat diproses lebih lanjut sehingga dihasilkan informasi yang lebih bermakna seperti rating.

Terdapat berbagai teknik klasifikasi yang dapat diterapkan pada sentimen analisis dalam mengklasifikasikan, salah satunya adalah Support Vector Machine (SVM). SVM berhasil menunjukkan hasil yang menjanjikan dalam berbagai penerapannya seperti pengenalan tulisan tangan dan kategorisasi teks. SVM dapat digunakan pada data yang berdimensi banyak dan menghindari kesulitan dari permasalahan dimensionalitas (Tan, 2006).

Penelitian mengenai teknik sentimen analisis dengan objek berupa tweet telah dilakukan sebelumnya oleh Saifinnuha, A. Z.

dengan judul “Penerapan Sentimen Analisis pada Twitter Berbahasa Indonesia untuk Mendapatkan Rating Program Televisi Menggunakan Metode Support Vector Machine”. Pada penelitiannya, Saifinnuha melakukan analisis terhadap sentimen yang diberikan oleh penonton acara televisi untuk menghasilkan rating yang didapat dari acara tersebut. Dari hasil penelitiannya, metode SVM dapat menghasilkan rating yang mendekati rating dari acara TV sebenarnya dengan rata-rata nilai error mencapai 0.59 (Saifinnuha, 2015).

Pada Penelitian lainnya dilakukan optimasi pemilihan fitur pada algoritma SVM dengan menggunakan algoritma genetika. Sistem melakukan klasifikasi teks kedalam opini positif atau negatif. Setelah penerapan algoritma genetika, ditunjukkan peningkatan hasil akurasi yang didapat, dari 71% menjadi 78.02% (Darmawan, 2015).

Pada penelitian ini, dengan mengambil judul “Penerapan Sentimen Analisis Acara Televisi pada Twitter Menggunakan Support Vector Machine dan Algoritma Genetika sebagai Metode Seleksi Fitur” akan menekankan pada penerapan seleksi fitur dengan Algoritma Genetika pada metode Support Vector Machine, untuk klasifikasi sentimen masyarakat mengenai suatu acara pada media Twitter. Melalui sentimen yang didapat, sistem akan melakukan perhitungan rating suatu acara berdasarkan jumlah sentimen positif dan negatif yang dihasilkan. Penerapan Algoritma Genetika dilakukan untuk melakukan seleksi fitur untuk dapat mengurai jumlah fitur yang digunakan namun tetap memperoleh hasil yang baik. Sehingga dengan menerapkan algoritma genetika, dapat diketahui perbandingan hasil yang diperoleh dari sentimen analisis dan perhitungan rating yang dihasilkan oleh sistem sebelum dan sesudah penerapan seleksi fitur.

2. DASAR TEORI

2.1 Rating Acara

Rating adalah tolak ukur yang dapat dijadikan acuan untuk melakukan evaluasi terhadap suatu produk. Menurut Supardi dalam (Saifinnuha, 2015), seorang produser acara TV akan mengambil keputusan apakah suatu program acara akan dilanjutkan atau tidak berdasarkan *rating* yang didapat, dari *rating*

yang didapat menunjukkan apakah program acara tersebut masuk kedalam kategori ditonton atau mulai kehilangan penontonnya.

Di Indonesia, beberapa stasiun TV nasional menggunakan jasa AGB Nielsen untuk melakukan perhitungan rating dari suatu program acara televisi (Supardi, 2012). AGB Nielsen Media Research melakukan perhitungan *rating* berdasarkan penonton yang ada pada 10 kota besar di Indonesia. Rating dari program acara didapat dengan persamaan 1. Untuk mengetahui *rating* dari suatu acara TV berdasarkan sentimen analisis didapatkan dengan mencari jumlah dari tweet yang mengandung sentimen positif lalu dimasukkan kedalam persamaan 2

$$Rating = \frac{Jumlah\ Penonton\ Program\ TV}{Total\ Penonton\ TV} \times 100\% \quad (1)$$

$$Rating_{acaraTV} = \frac{\sum Tweet\ Positif}{Populasi} \times 100\% \quad (2)$$

2.2 Sentimen analisis

Sentimen analisis yakni proses pembelajaran, memahami, mengekstran dan mengolah data yang berupa tekstual untuk mendapatkan informasi (Lee, 2008). Pada umumnya, sentimen analisis berupa klasifikasi data tekstual, namun pada kenyataannya sentimen analisis tidak mudah dilakukan dikarenakan perbedaan bahasa yang digunakan untuk menyusun isi dari suatu dokumen. Pada umumnya, teknik untuk melakukan sentimen analisis terbagi menjadi dua yaitu menggunakan teknik simbol, dan menggunakan teknik *machine learning*.

Teknik penggunaan simbol, dilakukan analisis terhadap setiap kata dari isi dokumen dan melakukan ekstraksi hubungan untuk mendapatkan sentimen. Pada teknik ini, perlu diketahui setia makna dan aturan kata dari tiap bahasa yang digunakan untuk menyusun isi dari dokumen dikarenakan tiap bahasa memiliki aturan dan makna yang berbeda-beda. Bahkan meskipun aturan baku dalam suatu bahasa telah dirumuskan sebelumnya, pada kenyataannya aturan tersebut tidak dapat diterapkan secara sempurna dikarenakan bahasa itu sendiri yang terus berkembang. Sedangkan, teknik *machine learning* menggunakan statistik untuk melakukan pemrosesan teks. Pada teknik ini, diaplikasikan ilmu maupun algoritma yang terdapat pada ilmu statistik seperti teori *bayes* atau teori populer lainnya yang dapat diterapkan pada proses *machine learning*.

2.3 Algoritma Genetika

Algoritma genetika adalah satu dari sekian metode yang dapat diterapkan pada SVM untuk mengatasi masalah seleksi fitur. Algoritma Genetika memiliki beberapa fase dalam menyelesaikan permasalahan yang dialami. Fase-fase tersebut dapat dijabarkan sebagai berikut.

1. Inisialisasi, merupakan proses untuk menciptakan individu-individu yang akan dijadikan solusi dari permasalahan yang ditangani.
2. Reproduksi, merupakan proses untuk menciptakan individu baru yang berasal dari individu sebelumnya. Proses reproduksi yang dilakukan dapat berupa *crossover* atau *mutation*. Hasil dari proses reproduksi akan ditempatkan kedalam kumpulan *offspring*.
3. Evaluasi, merupakan proses untuk menghitung nilai *fitness* dari individu *offspring* dan *parent*.
4. Seleksi, merupakan proses untuk memilih individu dengan nilai *fitness* terbaik dan diloloskan pada generasi selanjutnya (Mahmudy, 2015).

2.4 Sequential SVM

Metode SVM pertama kali diperkenalkan di tahun 1992 pada *Annual Workshop on Computational Learning Theory*. SVM dikembangkan oleh Boser bersama 2 orang temannya yakni Guyon dan Vapnik (Saifinnuha, A. Z., 2015). Prinsip kerja SVM pada awalnya sebagai metode untuk klasifikasi linier (*linear classifier*), dan dikembangkan untuk dapat menyelesaikan permasalahan klasifikasi non-linear, dengan memanfaatkan fungsi *kernel* untuk data atau ruang kerja dengan dimensi tinggi (Nugroho, 2007). Untuk dapat menyelesaikan permasalahan *non-linearly spareable*, data berdimensi D pada ruang vektor awal ($X_i \in R^D$) harus dipindahkan pada ruang vektor baru dengan dimensi yang lebih tinggi ($X'_i \in R^Q$) dengan fungsi pemetaan yang dapat dinotasikan dengan notasi $\Phi(x)$. Proses pemetaan vektor kedalam ruang vektor baru bertujuan untuk merepresentasikan data ke dalam yang *linearly spareable*. Selanjutnya dilakukan proses *training* seperti pada proses klasifikasi linear SVM. Proses optimasi pada fase *training* memerlukan nilai dari *dot product* yang didapat dari dua buah contoh pada ruang vektor baru. Nilai yang didapat dari

dot product vektor (X_i) dan (X_j) dinotasikan dengan $\Phi(X_i)$ dan $\Phi(X_j)$. Nilai *dot product* vektor (X_i) dan (X_j) dapat dihitung secara tidak langsung tanpa mengetahui fungsi Φ , dengan memanfaatkan fungsi *kernel*. Teknik komputasi *kernel*, yaitu menghitung *dot product* dua vektor pada ruang vektor baru dengan memakai komponen kedua buah vektor tersebut di ruang vektor asal. Persamaan *kernel* dapat dilihat pada persamaan 3. Tabel 1 berisikan berbagai fungsi *kernel* yang umum digunakan.

$$(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \tag{3}$$

Tabel 1. Fungsi *Kernel*

Nama Kernel	Fungsi
Polynomial	$K(x_i, x_j) = (x_i \cdot x_j + 1)^p$
Gaussian (RBF)	$K(x_i, x_j) = \exp\left(-\left(\frac{ x_i, x_j ^2}{2\sigma^2}\right)\right)$

Pada tahun 1999, Vijayakumar & Wu mengembangkan *sequential training* SVM untuk mendapatkan *hyperplane*. Metode yang diusulkan memberikan solusi yang optimal dengan mempercepat proses iterasi serta menghasilkan urutan yang lebih cepat daripada metode konvensional. Langkah-langkah *sequential training*, dapat diurutkan sebagai berikut:

1. Inisialisasi parameter yang digunakan pada tahap *sequential training* yakni λ (variabel skalar), γ (*learning rate*), C (variabel *slack*), ϵ (epsilon) dan iterasi maksimum.
2. Lanjutkan dengan inisialisasi $\alpha_i = 0$ lalu hitung nilai Matrik Hessian D_{ij} dengan persamaan 4. Dimana y adalah kelas data, x adalah nilai dari data dan K adalah kernel yang digunakan.

$$D_{ij} = y_i y_j (K(x_i, x_j) + \lambda^2) \tag{4}$$
3. Untuk seluruh data, ulangi langkah a), b) dan c) berikut untuk memperbaharui nilai E (*error*) dan α (*alpha*).
 - a) $E_i = \sum_i^n \alpha_i D_{ij} \tag{5}$
 - b) $\delta\alpha = \min\{\max[\gamma(1 - E_i), -\alpha_i], C - \alpha_i\} \tag{6}$
 - c) $\alpha_i = \alpha_i + \delta\alpha_i \tag{7}$
4. Ulangi langkah ketiga hingga mencapai batas iterasi maksimum atau nilai dari $Max(|\delta\alpha|) < \epsilon$.
5. Selesai

Dari proses 1-5 dokumen yang memiliki nilai $\alpha > 0$ disebut dengan *support vector*. Lalu menentukan *support vector* dengan nilai α tertinggi lalu dilambangkan dengan X^+ (kelas positif) dan X^- (kelas negatif). Selanjutnya menghitung nilai bias b dengan persamaan 8 dengan menggunakan kernel yang sudah ditentukan.

$$b = -\frac{1}{2}(\sum_{i=0}^n a_i y_i K(X_i, x^-) + \sum_{i=0}^n a_i y_i K(X_i, x^+)) \tag{8}$$

Sentimen analisis dapat dihitung dengan persamaan 9. Dimana jika nilai $SIGN(h(x))$ menunjukkan +1 maka termasuk kelas positif dan nilai -1 untuk kelas negatif.

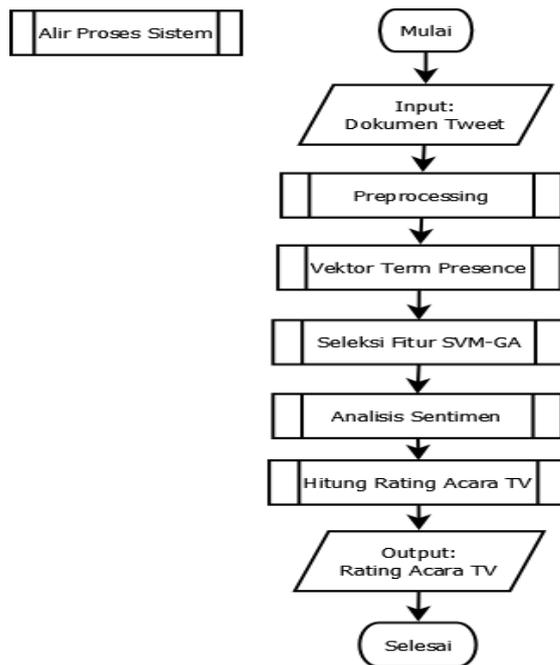
$$h(x) = \sum_{i=0}^n a_i y_i K(x, x_i) + b \tag{9}$$

Dimana b adalah nilai bias yang diperoleh pada proses *training*, y adalah kelas data *training*, x_i adalah nilai dari data *training*, dan x adalah data *testing*.

3. PERANCANGAN & IMPLEMENTASI

3.1. Alur Proses Sistem

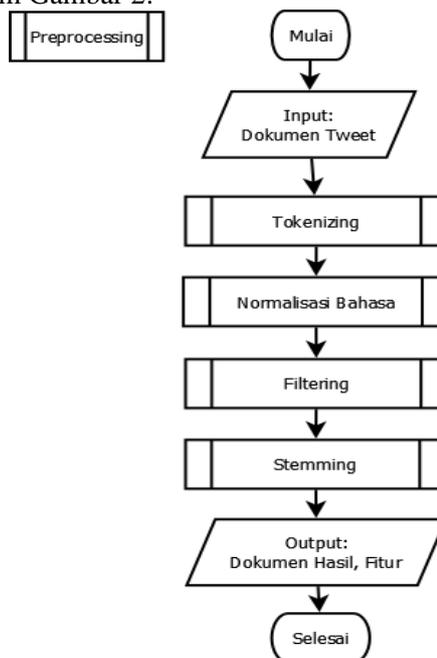
Alur proses sentimen analisis untuk mendapatkan rating acara tv dibagi kedalam beberapa proses. Urutan proses pada sistem adalah, memasukkan dokumen tweet, *preprocessing*, *vector term presence*, seleksi fitur SVM-GA, sentimen analisis, lalu perhitungan *rating* (Saifinnuha, 2015). Gambar 1 menunjukkan diagram alir dari sistem.



Gambar 1. Alir Proses Sistem

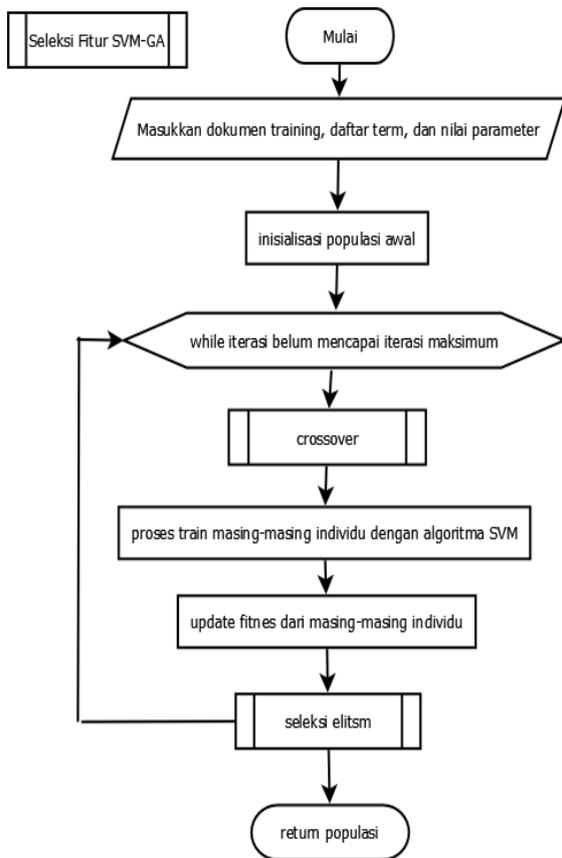
3.2 Preprocessing

Preprocessing adalah proses awal untuk mempersiapkan dokumen yang akan digunakan sebagai *input* pada proses selanjutnya. Tahap *preprocessing* terbagi menjadi beberapa proses yakni, *tokenizing*, normalisasi bahasa, *filtering*, dan *stemming*. Diagram alir proses *preprocessing* ditunjukkan dalam Gambar 2.



Gambar 2. Alir Preprocessing

3.2 Seleksi Fitur SVM-GA



Gambar 3. Seleksi Fitur SVM-GA

Proses seleksi fitur bertujuan untuk mengurangi jumlah *term* yang didapat dari proses *preprocessing* yang nantinya digunakan pada proses sentimen analisis. Alir proses seleksi fitur dapat dilihat dalam Gambar 3. Proses dimulai dengan membentuk individu awal yang memiliki *chromosome* berupa kombinasi *term* yang dipilih secara acak. Individu tersebut lalu dijadikan sebagai *parent* pada proses reproduksi, untuk mendapatkan *offspring* baru. Selanjutnya, individu yang terbentuk pada proses inisialisasi dan reproduksi dimasukkan kedalam proses *training* dengan menggunakan metode SVM untuk mendapatkan *fitness* dari masing-masing individu. Individu dengan nilai *fitness* yang baik, akan lolos dari proses seleksi dan masuk kedalam proses iterasi selanjutnya.

4. PENGUJIAN DAN ANALISIS

Pengujian dilakukan untuk mengetahui pengaruh peningkatan jumlah data *training* dan data *testing*, jenis kernel, ukuran *popsize* dan nilai *crossover* rate yang digunakan. Pengujian pada parameter lain tidak dilakukan, karena

berpengaruh terhadap proses pengolahan data yang semakin lama (Darmawan, 2015).

4.1 Hasil Pengujian Rasio Data *Training*, dan Data *Testing*, dan Pengaruh Kernel

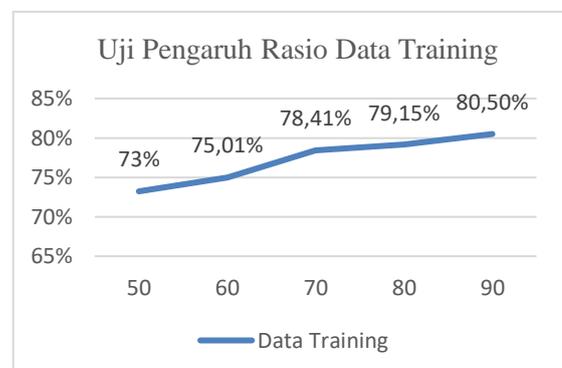
Pada pengujian pengaruh rasio data *training*, dan data *testing*, dan kernel terhadap akurasi, digunakan parameter yang sudah ditetapkan sebagai berikut $\lambda = 0.5$, γ (*learning rate*) = 0.001, C (*variabel slack*) = 1, ϵ (*epsilon*) = 0.0001, Iterasi maksimum SVM = 50, *popsize* = 5, *crossover rate* = 0.2, dan iterasi maksimum algen = 5.

Jumlah data yang digunakan dalam pengujian adalah sebanyak 160 tweet yang terbagi menjadi 80 tweet bersentimen positif dan 80 tweet bersentimen negatif. Berdasarkan perancangan pengujian rasio data *training* dan data *testing*, terdapat 5 rasio perbandingan yang akan diuji. Rasio perbandingan tertera pada Tabel 2.

Tabel 2. Rasio Perbandingan Data *Training* dan Data *Testing*

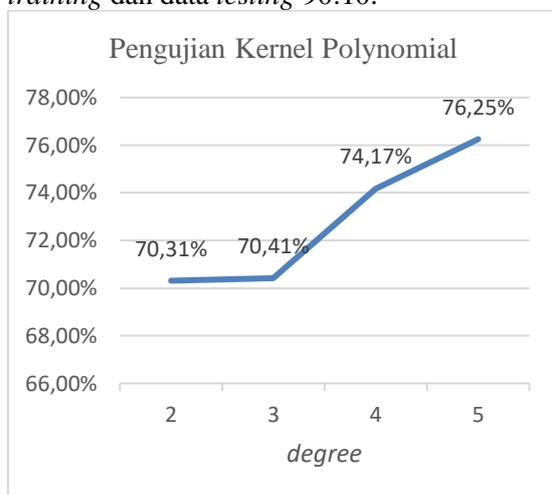
Rasio Perbandingan	Jumlah Data <i>Training</i>	Jumlah Data <i>Testing</i>
90:10	144	16
80:20	128	32
70:30	112	48
60:40	96	64
50:50	80	80

Masing-masing rasio akan diberikan 9 skenario pengujian dengan menggunakan kernel yang berbeda. Masing-masing skenario akan dilakukan pengujian sebanyak 5 kali untuk mendapatkan rata-rata dari akurasi.

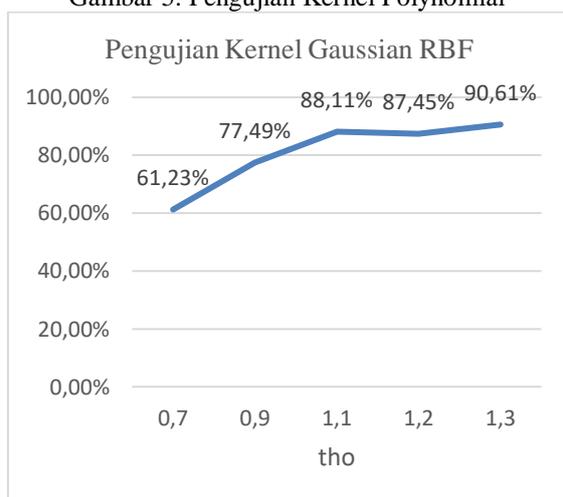


Gambar 4. Pengaruh Jumlah Data *Training*

Berdasarkan Gambar 4, ditunjukkan pengaruh dari rasio perbandingan data *training* dan data *testing* terhadap hasil akurasi. Akurasi terbesar didapatkan pada rasio data *training* 90%. Seiring dengan peningkatan jumlah data *training*, akurasi yang didapatkan terus meningkat dikarenakan bertambahnya jumlah *term* yang dapat digunakan. Hal ini disebabkan karena, semakin banyaknya pengetahuan yang dimiliki oleh sistem. Rasio dengan hasil akurasi terbaik yang didapat digunakan untuk pengujian perbandingan kernel terhadap akurasi. Pada Gambar 5 dan Gambar 6 ditunjukkan perbandingan akurasi untuk masing-masing kernel dengan rasio data *training* dan data *testing* 90:10.



Gambar 5. Pengujian Kernel Polynomial



Gambar 6. Pengujian Kernel Gaussian RBF

Berdasarkan Gambar 5 dan Gambar 6 didapatkan bahwa kernel Gaussian RBF menghasilkan rata-rata akurasi yang lebih baik daripada kernel polynomial dengan akurasi tertinggi berada pada nilai θ 1.3 yakni 90.61%. Peningkatan nilai parameter yang

digunakan pada masing-masing kernel menunjukkan terjadinya peningkatan pada hasil akurasi yang diperoleh. Pada kernel *polynomial* akurasi terbaik dihasilkan ketika nilai variabel *degree* yang digunakan adalah 5 dengan akurasi yang diperoleh sebesar 76.25%, sedangkan untuk kernel Gaussian RBF, akurasi terbaik adalah 90.61%, ketika θ bernilai 1.3.

4.2 Hasil Pengujian Pengaruh Pengetahuan Data Training

Pengujian dilakukan dengan menggunakan data *testing* sebanyak 64 dokumen tweet untuk tiap jumlah data *training* yang diuji. Masing-masing rasio data *training* dilakukan pengujian sebanyak 5 kali untuk mendapatkan rata-rata akurasi yang diperoleh.



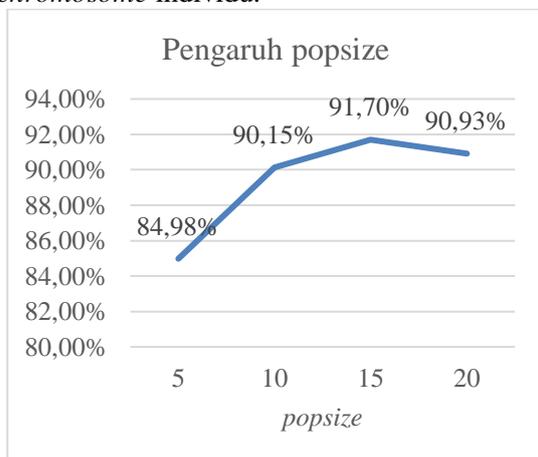
Gambar 7. Pengujian Pengaruh Pengetahuan Data Training

Berdasarkan Gambar 7 ditunjukkan pengaruh dari peningkatan jumlah data *training*, terhadap akurasi yang diperoleh. Akurasi terendah dihasilkan ketika jumlah data *training* yang dipergunakan sebanyak 96 dokumen dan akurasi tertinggi dihasilkan ketika jumlah data *training* sebanyak 144 dokumen. Hal ini disebabkan karena semakin banyak jumlah dokumen yang dipergunakan sebagai data *training*, sistem akan memiliki pengetahuan semakin banyak mengenai ragam *term* yang dapat digunakan. Semakin tingginya pengetahuan yang dimiliki sistem, akan dapat menghasilkan nilai akurasi yang lebih baik dalam melakukan proses klasifikasi.

4.3 Hasil Pengujian Pengaruh Parameter *popsize* dan *crossover rate*

Pada pengujian pengaruh parameter *popsize* dan *crossover rate* terhadap hasil akurasi akan digunakan perbandingan rasio 90:10 untuk jumlah data *training* dan data *testing* dan kernel gaussian RBF dengan parameter σ 1.3. Jumlah iterasi maksimum yang dilakukan untuk proses SVM adalah 50 dan untuk proses algen adalah 5. Terdapat 4 skenario untuk nilai yang akan digunakan sebagai parameter *popsize* dengan 4 skenario untuk nilai *crossover rate*. Masing-masing skenario akan dilakukan pengujian sebanyak 5 kali untuk mendapatkan rata-rata akurasi terbaik.

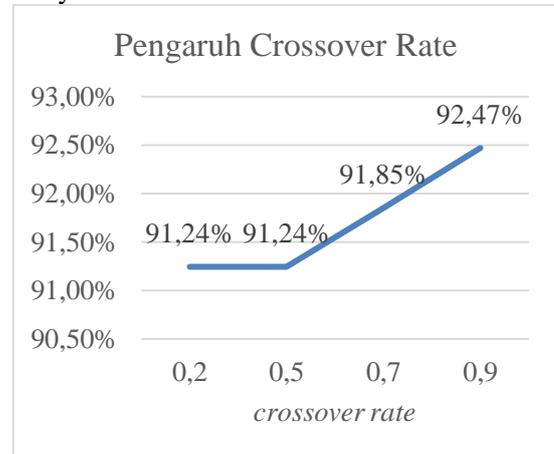
Berdasarkan Gambar 8 didapatkan bahwa jumlah *popsize* dengan rata-rata akurasi tertinggi berada pada nilai 15 dengan nilai akurasi 91.701%. Peningkatan nilai *popsize* dari bernilai 5 hingga 15 menunjukkan akurasi yang bagus, dikarenakan semakin banyaknya jumlah individu yang digunakan maka semakin banyak pula kombinasi *term* yang dapat dipilih. Namun, terjadi penurunan pada saat *popsize* bernilai 20, hal ini disebabkan karena tidak munculnya *term* yang dapat berpengaruh besar terhadap akurasi, didalam salah satu *chromosome* individu.



Gambar 8. Hasil Pengujian *Popsize*

Pada Gambar 9 ditampilkan akurasi yang didapatkan dengan meningkatkan nilai dari parameter *crossover rate*. Ditunjukkan bahwa dengan peningkatan nilai *crossover rate*, nilai akurasi yang diperoleh juga ikut meningkat. Hal ini disebabkan karena semakin banyakna *offspring* yang dihasilkan maka semakin banyak kombinasi *term* yang didapat sehingga memungkinkan untuk *term* yang

memiliki pengaruh besar terhadap proses klasifikasi untuk ikut terpilih. Akurasi tertinggi ditunjukkan ketika nilai *crossover rate* adalah 0.9 yakni 92.47%.



Gambar 9. Hasil Pengujian *crossover rate*

4.4 Hasil Pengujian Perbandingan Akurasi Sebelum dan Sesudah Diterapkan Seleksi Fitur

Pada pengujian perbandingan akurasi sebelum dan sesudah diterapkan seleksi fitur akan digunakan perbandingan rasio 90:10 untuk jumlah data *training* dan data *testing* dan kernel gaussian RBF dengan parameter σ 1.3. Pada parameter algen, untuk *popsize* sebesar 15 dan *crossover rate* 0.9. Jumlah iterasi maksimum yang dilakukan untuk proses SVM adalah 50 dan 5 untuk proses algen.

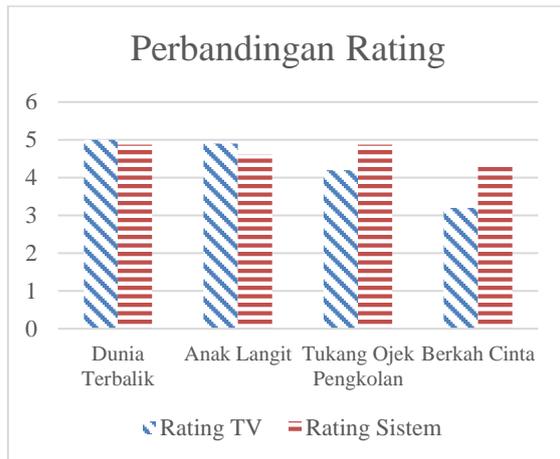


Gambar 10. Hasil Perbandingan Akurasi Sebelum dan Sesudah Diterapkan Seleksi Fitur

Pada Gambar 10 ditunjukkan perbandingan akurasi yang didapatkan oleh sistem, sebelum dan sesudah menerapkan seleksi fitur. Dari hasil yang didapat, dengan mengurangi jumlah fitur yang digunakan secara acak, sistem dapat mendapatkan akurasi

yang hamper menyerupai kondisi sistem sebelum menerapkan seleksi fitur dengan nilai rata-rata nilai *error* 0.62%. Dari lima kali percobaan pengujian yang dilakukan, sistem hanya mendapatkan satu kali hasil akurasi yang berbeda dengan kondisi sebelum menerapkan seleksi fitur, yakni pada percobaan pertama dengan nilai akurasi 87.50%.

4.5 Hasil Pengujian *Rating* Program TV



Gambar 11. Hasil Perbandingan *Rating*

Pada pengujian perbandingan *rating* Program TV yang diperoleh dari sistem dengan *rating* sesungguhnya, digunakan perbandingan rasio 90:10 untuk jumlah data *training* dan data *testing* dan kernel gaussian RBF dengan parameter σ 1.3. Pada parameter algen, untuk *popsize* sebesar 15 dan *crossover rate* 0.9. Jumlah iterasi maksimum yang dilakukan untuk proses SVM adalah 50 dan 5 untuk proses algen.

Pada Gambar 11 ditunjukkan perbandingan *rating* dari TV dengan *rating* dari sistem. *Rating* yang dihasilkan oleh sistem menunjukkan hasil yang mendekati *rating* sebenarnya, dengan rata-rata nilai *error* 0,562. Nilai *error* terkecil didapat pada *rating* acara Dunia Terbalik dimana sistem memberikan *rating* sebesar 4,875 sedangkan *rating* sebenarnya adalah 5 sehingga didapatkan nilai *error* senilai 0,125. Dan nilai *error* terbesar terdapat pada acara Berkah Cinta yakni senilai 1,175 dimana sistem memberikan *rating* sebesar 4,375 sedangkan *rating* sebenarnya adalah 3,2. Berdasarkan urutan dari *rating* terbesar hingga terkecil, sistem mampu mengurutkan acara dengan hasil 1 kesalahan. Pada Tabel 6.7, dengan mengurutkan hasil *rating* terbesar hingga *rating* terkecil yang

didapat dari masing-masing acara, maka dapat diketahui peringkat dari masing-masing acara. Sistem dapat memberikan peringkat yang tepat untuk masing-masing acara dengan satu kesalahan pada peringkat 2 dan 3.

5. KESIMPULAN

Berdasarkan hasil penelitian tentang penerapan sistem analisis acara televisi pada twitter menggunakan *support vector machine* dan algoritma genetika sebagai seleksi fitur, bisa disimpulkan bahwa, metode *support vector machine* dan algoritma genetika dapat diterapkan untuk melakukan sentimen analisis opini masyarakat Indonesia yang berupa tweet, untuk mendapatkan *rating* dari program TV. Data Tweet yang dipergunakan melalui proses *preprocessing* (*tokenizing*, normalisasi bahasa, *filtering*, dan *stemming*), *vector term presence*, dan seleksi SVM-GA. Hasil akhir dari proses seleksi SVM-GA berupa individu dengan kombinasi *term* dan akurasi terbaik yang akan digunakan untuk proses *testing* sentimen analisis. Sistem menghitung *rating* dari suatu program acara TV dengan mengumpulkan seluruh tweet beropini positif mengenai program acara tersebut lalu dibagi dengan populasi seluruh tweet. Berdasarkan hasil implementasi, *rating* yang didapat dari penerapan sentimen analisis acara televisi pada Twitter dapat dipergunakan untuk menunjukkan *rating* sebenarnya dari suatu program acara TV. *Rating* dihasilkan dengan mengumpulkan seluruh tweet positif yang tertuju pada suatu acara kemudian dibagi dengan jumlah populasi tweet yang dipergunakan. Sistem mampu mencapai rata-rata nilai *error* yang cukup kecil yaitu 0,562 untuk rentang nilai 0-5.

Berdasarkan hasil pengujian yang dilakukan, algoritma genetika dapat diterapkan untuk melakukan seleksi fitur yang akan digunakan pada metode *support vector machine* dalam sentimen analisis. Dengan diterapkan seleksi fitur, sistem dapat menggunakan jumlah fitur yang lebih sedikit namun memperoleh hasil akurasi yang mendekati kondisi menggunakan seluruh fitur. Dari hasil penerapan, sistem mendapatkan nilai *error* 0.62 untuk perbedaan akurasi yang didapat sebelum dan sesudah diterapkan seleksi fitur.

DAFTAR PUSTAKA

- Capece, D. K. L. E. & L. T., 2009. *Digital Influence in News and Politics*. New York: Sparxoo.
- Darmawan, A., 2015. Penerapan Model Support Vector Machine Text Mining Pada Komentar Review Smartphone Android VS Blackberry Dengan Teknik Optimasi Genetic Algorithm. *Faktor Exacta*, Volume 8, pp. 100-115.
- Fuvenanda, 2010. *Twitter dan Pengaruhnya Terhadap Perkembangan Informasi*. [Online] Available at: <http://fuvenanda.wordpress.com> [Accessed 7 Maret 2017].
- Lee, B. P. A. L., 2008. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, Volume 2, pp. 1-135.
- Mahmudy, W., 2015. *Modul Kuliah Dasar-Dasar Algoritma Evolusi*. s.l.:Universitas Brawijaya.
- MIT, 2013. *Twitter`s World*. [Online] Available at: <https://www.technologyreviews.com> [Accessed 7 Maret 2017].
- Nugroho, A., 2007. *Pengantar Support Vector Machine*. s.l.:Chukyo University.
- Patil, G., 2014. Sentiment Analysis Using Support Vector Machine. *International Journal of Inovative Research in Computer and Communication Engineering*, pp. 2607-2612.
- Saifinnuha, A., 2015. Penerapan Sentimen Analisis pada Twitter Berbahasa Indonesia untuk mendapatkan Rating Program Televisi Menggunakan Metode Support Vector Machine. *Universitas Brawijaya*.
- Supardi, F. D., 2012. *Mekanisme Perhitungan Rating TV Di Indonesia*. [Online] Available at: <http://ferydedi.blogspot.co.id/2012/08/mekanisme-perhitungan-rating-tv-indonesia.html> [Accessed 7 Maret 2017].
- Tan, P. S. M. & K. V., 2006. *Introduction to Data Mining*. Addison Wesley ed. Boston: Pearson.