

## ANALISIS SENTIMEN TERHADAP TOKOH PUBLIK MENGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE* (SVM)

I. Taufik<sup>1</sup> dan S.A. Pamungkas<sup>2</sup>

<sup>1</sup>Program Studi Matematika, Fakultas Sains dan Teknologi  
Universitas Islam Negeri Syarif Hidayatullah Jakarta

<sup>2</sup>Badan Pengkajian dan Penerapan Teknologi (BPPT) Jakarta  
Email:

**Abstract:** Sentiment analysis is a computational study of people's opinions. The sentiment analysis will cluster text in sentences or documents to find out the opinions in these sentences or documents, either positive or negative. In this research, we discuss sentiment analysis for public figure on Twitter. We used data tweet with keywords "Ahok" and "@teman\_ahok". We processed the tweets and then classified it using the support vector machine algorithm. There are four kernel parameters in this algorithm: linear kernel, sigmoid kernel, gaussian kernel and polynomial kernel. The performance of this algorithm will be seen using precision, accuracy, and recall. The result shows that the linear kernel has the best level of precision is about 80%. The sigmoid kernel has the best recall rate of about 85%. The sigmoid kernel has the best accuracy rate of about 81%.

**Keywords:** Sentiment Analysis, Support Vector Machines, Text Mining.

**Abstrak:** *Sentiment analysis* atau analisis sentimen adalah studi komputasional dari opini-opini orang. Analisis sentimen akan mengelompokkan teks yang ada dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam kalimat atau dokumen tersebut, bisa berupa positif atau negatif. Skripsi ini membahas tentang sentimen seseorang di Twitter terhadap tokoh publik. Data yang digunakan berupa data *tweet* dengan kata kunci "Ahok" dan "@teman\_ahok". *Tweet* yang didapat kemudian diolah dengan melakukan *text preprocessing* kemudian diklasifikasikan dengan menggunakan algoritma *support vector machine*. Ada empat parameter kernel dalam algoritma ini: kernel linear, kernel sigmoid, kernel gaussian, dan kernel polinomial. Kinerja algoritma ini menggunakan presisi, akurasi, dan *recall*. Hasil menunjukkan bahwa *Kernel linear* memiliki tingkat presisi yang paling baik sekitar 80%. *Kernel sigmoid* memiliki tingkat *recall* yang paling baik sekitar 85%. *Kernel sigmoid* memiliki tingkat akurasi yang paling baik sekitar 81%.

**Kata kunci:** *Sentiment Analysis, Support Vector Machines, Text Mining.*

### PENDAHULUAN

Pada saat ini situs *microblogging* telah menjadi alat komunikasi yang sangat populer di kalangan pengguna internet. Dimana jutaan pesan yang muncul setiap hari di situs web populer yang menyediakan layanan *microblogging* seperti Twitter, Tumblr dan Facebook. Penulis pesan tersebut menulis tentang kehidupan mereka, berbagai opini tentang berbagai topik dan membahas isu-isu yang terjadi saat ini. Format pesan yang bebas dan aksesibilitas dari berbagai *platform* yang mudah, pengguna internet cenderung untuk beralih dari blog atau milis ke layanan *microblogging* [1].

Twitter merupakan salah satu situs *microblogging* dengan pengguna lebih dari 29,5 juta di Indonesia dan 383 juta *tweet* per hari [10]. Twitter memungkinkan penggunanya untuk berbagi pesan menggunakan teks pendek yang disebut *tweet*. Twitter dapat menjadi sumber data pendapat dan sentimen masyarakat yang dapat digunakan secara efisien untuk pemasaran atau studi sosial [8].

Studi komputasional dari opini-opini pengguna internet, sentimen dan emosi melalui entitas dan atribut yang dimiliki dan diekspresikan dalam bentuk teks disebut dengan *Sentiment Analysis* atau *Opinion Mining* [7]. Analisis sentimen akan mengelompokkan polaritas dari teks ke dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam bentuk kalimat atau dokumen tersebut bersifat positif, negatif atau netral [9].

Beberapa penelitian telah dilakukan terkait *sentimen analysis*. Aliandu [2] telah melakukan sentimen publik yang disampaikan melalui tweet berbahasa Indonesia. Penelitian ini menggunakan Algoritma *Naive Bayes* dan juga menggunakan *emoticon* untuk memudahkan anotasi terhadap kelas sentimen dari *data training*. Romelta [9] menggunakan dua buah algoritma pembelajaran yaitu *Support Vector Machine* dan *Naive Bayes* untuk menggali opini Customs terhadap produk *smartphone* di Twitter. Pada penelitian lain, Vidya dkk. [10] melakukan *sentimen analysis* terhadap reputasi *provider handphone* dengan menggunakan SVM, *Naive Bayes* dan *Decision Tree*.

Pada penelitian ini akan dilakukan *sentimen analysis* terhadap tokoh publik dengan menggunakan algoritma *Support Vector Machine* (SVM). Data diambil dari Twitter yang dipisahkan menjadi *data training* dan *data testing*. Data training digunakan untuk memperoleh parameter-parameter yang digunakan pada proses pengklasifikasian menggunakan Algoritma SVM sedangkan data testing digunakan untuk mengevaluasi hasil pengklasifikasian. Interpretasi penelitian ini hanyalah semata-mata berdasarkan metode dan tidak dipengaruhi oleh opini-opini dan tujuan politik lainnya.

## METODOLOGI PENELITIAN

### Data Penelitian

Data yang digunakan dalam penelitian ini diambil dari Twitter dari salah satu tokoh publik dengan kata kunci “@temanAhok” dan “Ahok” dan data *stopword*. Data Twitter diambil dari tanggal 1-30 Mei 2016, sedangkan data *stopword* diambil dari internet. Data *stopword* terdiri dari 1490 kata.

### Pengolahan Data

Pengolahan data dilakukan melalui proses berikut:

1. Proses preprocessing yang dilakukan pada penelitian ini melalui 4 tahapan yaitu case folding, tokenizing, filtering dan stemming. Tahapan preprocessing akan menghasilkan kumpulan atribut atau keyword dalam bentuk term (kata per kata).
2. Tahap tokenizing. Pada tahap ini, karakter-karakter tertentu seperti tanda baca dihilangkan. Sedangkan karakter spasi digunakan sebagai delimiter untuk memecah kalimat menjadi kumpulan kata.
3. Proses filtering menggunakan stop-list dengan 532 daftar kata *stopword*. Jika dalam suatu data terdapat *stopword*, maka secara otomatis kata-kata tersebut akan dihapus. Kemudian kata-kata yang terdapat dalam proses filtering diubah ke dalam bentuk kata dasar dengan

menggunakan aturan tertentu (stemming). Pada penelitian ini, proses stemming dilakukan menggunakan Algoritma Porter.

4. Pembobotan kata. Misalkan terdapat sekumpulan *term*  $t_j = (t_1, t_2, \dots, t_m)$  dan dokumen  $d_i = (d_1, d_2, \dots, d_n)$ . Pembobotan kata dilakukan dengan rumus berikut :

$$W_{ij} = tf_{ij} \cdot idf$$

$$W_{ij} = tf_{ij} \cdot \log \left( \frac{n}{df} \right)$$

dengan  $W_{ij}$  bobot *term* ( $t_j$ ) terhadap dokumen ( $d_i$ ),  $tf_{ij}$  jumlah kemunculan *term* ( $t_j$ ) dalam ( $d_i$ ),  $n$  jumlah semua dokumen yang ada dalam *database*, dan  $df$  banyaknya dokumen yang mengandung *term*.

5. Normalisasi. Normalisasi adalah sebuah cara untuk menormalkan vektor dokumen sehingga proses tidak terpengaruh oleh perbedaan panjang dokumen. Proses normalisasi panjang dokumen tersebut dinamakan Cosine Normalization. Sehingga bobot normalisasi panjang dokumen untuk setiap *term* mempunyai nilai antara 0-1. Normalisasi didapatkan dengan rumus:

$$x_{ij} = \frac{W_{ij}}{\sqrt{W_{i1}^2 + W_{i2}^2 + \dots + W_{im}^2}} \quad (1)$$

dengan  $x_{ij}$  bobot hasil normalisasi *term* ( $t_j$ ) terhadap dokumen ( $d_i$ ) dan  $W_{ij}$  bobot *term* ( $t_j$ ) terhadap dokumen ( $d_i$ ).

6. Klasifikasi menggunakan Algoritma *Support Vector Machine* (SVM). Cara kerja SVM adalah sebagai berikut:
  - a. Pembagian data menjadi data latih dan data uji. Pembagian data dilakukan secara acak. Data pertama sampai ke-420 dijadikan data latih sedangkan data ke-421 sampai data ke-630 dijadikan sebagai data uji. Data latih digunakan untuk proses pelatihan dalam mencari nilai  $\alpha$  dan  $b$  untuk digunakan pada fungsi keputusan (*classifier*). Sedangkan data uji digunakan untuk proses pengujian yang nantinya akan menghasilkan klasifikasi data apakah positif ataukah negatif. Data yang menyatakan tweet positif diberi label 1 sedangkan data yang menyatakan tweet negatif diberi label -1.
  - b. Kernelisasi data latih dengan menggunakan fungsi kernel linear  $K(x_i, x_j) = (x_i, x_j)$ . Pemilihan kernel linear dilakukan karena kernel linear adalah kernel terbaik untuk data teks. Fungsi kernel diatas akan menghasilkan matriks kernel. Setiap elemen matriks kernel  $K(x_i, x_j)$  digunakan untuk menggantikan *dot-product*  $x_i \cdot x_j$  dalam persamaan dualitas Lagrange:

$$\begin{aligned} \max L_D(\alpha) &= \frac{1}{2} \left( \sum_{i=1}^n \alpha_i y_i x_i \right) \left( \sum_{i=1}^n \alpha_i y_i x_i \right) - \sum_{i=1}^n \alpha_i y_i \left( \left( \sum_{i=1}^n \alpha_i y_i x_i \right) x_i + b \right) + \sum_{i=1}^n \alpha_i \\ &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i y_i \alpha_j y_j (x_i \cdot x_j) - \sum_{i=1}^n \sum_{j=1}^n \alpha_i y_i \alpha_j y_j (x_i \cdot x_j) - b \sum_{i=1}^n \alpha_i y_i + \sum_{i=1}^n \alpha_i \\ &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i y_i \alpha_j y_j (x_i \cdot x_j) \end{aligned}$$

dengan kendala

$$\sum_{i=1}^n \alpha_i y_i = 0, \quad \alpha_i \geq 0, i = 1, 2, \dots, n$$

- c. Dari proses pelatihan diperoleh nilai parameter  $\alpha$  dan  $b$ .
- d. Setelah proses pelatihan selesai maka selanjutnya proses pengujian untuk melihat tingkat akurasi dan kesalahan dari sistem tersebut. kesalahan dari sistem tersebut. Umumnya, pengukuran kinerja klasifikasi dilakukan dengan matriks konfusi (*confussion matriks*). Matriks konfusi merupakan tabel pencatat hasil kerja klasifikasi. Tabel 1 merupakan contoh matriks konfusi yang melakukan klasifikasi masalah biner pada dua kelas yaitu kelas 0 dan kelas 1. Setiap sel  $f_{ij}$  dalam matriks menyatakan jumlah *record* (data) dari kelas  $i$  yang hasil prediksinya masuk ke kelas  $j$ . Misalnya sel  $f_{11}$  adalah jumlah data dalam kelas 1 yang secara benar dipetakan ke kelas 1 dan  $f_{10}$  adalah data dalam kelas 1 yang dipetakan secara salah ke kelas 0.

**Tabel 1.** Contoh Matriks Konfusi

$f_{ij}$		Kelas hasil prediksi ( $j$ )	
		Kelas = 1	Kelas = 0
Kelas asli ( $i$ )	Kelas = 1	$f_{11}$ ( <i>True Positive</i> )	$f_{10}$ ( <i>False Negative</i> )
	Kelas = 0	$f_{01}$ ( <i>False Positive</i> )	$f_{00}$ ( <i>True Negative</i> )

- e. Berdasarkan isi matriks konfusi, dapat diketahui nilai presisi, recall dan akurasi dari hasil klasifikasi yang diperoleh. Untuk menghitung akurasi, presisi dan recall digunakan formula:

$$\begin{aligned} \text{Presisi} &= \frac{f_{11}}{f_{11} + f_{01}}, \\ \text{Recall} &= \frac{f_{11}}{f_{11} + f_{10}}, \\ \text{Akurasi} &= \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}}. \end{aligned}$$

## HASIL DAN PEMBAHASAN

Data *training* pada penelitian ini berjumlah 420 data yang telah dilabeli sentimennya oleh penulis dan digunakan untuk proses *training* algoritma. Tabel 2 adalah contoh data *training* yang telah dilabeli sentimennya oleh penulis.

**Tabel 2.** Contoh Data *Training*

No	<i>Tweet</i>	Label Menurut Peneliti
1	@Ronni_Hermawan @basuki_btp @kangdede78 Alhamdulillah warga Bekasi bisa ke Jakarta pp dgn bis TJ hanya dgn 3500. Terimakasih Ahok	Positif
2	@Reiza_Patters @temanAhok Kalo gue bilang Ahok bego kalian mau apa? Gue punya KTP DKI hak gue buat nilai Gubernur bego dan gagal.	Negatif
3	@RatnaSpaet @SiBonekaKayu @basuki_btp @Aktualco itulah tipu daya ahok sungguh licik dan munafik...	Negatif

## Analisis Sentimen terhadap Tokoh Publik menggunakan Algoritma *Support Vector Machine*

Data testing yang digunakan pada penelitian ini berjumlah 210 dan telah dilabeli sentimennya. Ini dilakukan untuk mengevaluasi kinerja sistem. Tabel 3 adalah contoh data testing.

**Tabel 3.** Contoh *Data Testing*

No	<i>Tweet</i>
1	INILAH AHOK @basuki_btp MUSANG BERBULU DOMBA \nKAFIR BERBUSANA MUSLIM. pic.twitter.com/v8bdLXkZGC

Pada *preprocessing* text meliputi tahapan-tahapan *case folding* dijelaskan pada Tabel 4, *Tokenizing* pada Tabel 5, *Filtering* pada Tabel 6 dan *Stemming* pada Tabel 7.

**Tabel 4.** Proses *case folding*

No	<i>Tweet</i>	Casefolding
1	@Ronni_Hermawan_@basuki_btp @kangdede78 Alhamdulillah warga Bekasi bisa ke Jakarta pp dgn bis TJ hanya dgn 3500. Terima kasih Ahok	@ronni_hermawan_@basuki_btp @kangdede78 alhamdulillah warga beklasi bisa ke jakarta pp dgn bis tj hanya dgn 3500. terima kasih ahok
2	@Reiza_Patters @temanAhok Kalo gue bilang Ahok bego kalian mau apa? Gue punya KTP DKI hak gue buat nilai Gubernur bego dan gagal.	@reiza_patters @temanahok kalo gue bilang ahok bego kalian mau apa? gue punya ktp dki hak gue buat nilai gubernur bego dan gagal.
3	@RatnaSpaet @SiBonekaKayu @basuki_btp @Aktualco itulah tipu daya ahok sungguh licik dan munafik...	@ratnaspamet @sibonekakayu @basuki_btp @aktualco itulah tipu daya ahok sungguh licik dan munafik...
4	INILAH AHOK @basuki_btp MUSANG BERBULU DOMBA \nKAFIR BERBUSANA MUSLIM. pic.twitter.com/v8bdLXkZGC	inilah ahok @basuki_btp musang berbulu domba \nkafir berbusana muslim. pic.twitter.com/v8bdlxkzgc

**Tabel 5.** Proses *Tokenizing*

No	<i>Tweet</i>	<i>Tokenizing</i>
1	@Ronni_Hermawan_@basuki_btp @kangdede78 Alhamdulillah warga Bekasi bisa ke Jakarta pp dgn bis TJ hanya dgn 3500. Terima kasih Ahok	ronnihermawan basukibtp kangdede78 alhamdulillah warga bekasi bisa ke jakarta pp dgn bis tj hanya dgn 3500 terima kasih ahok

## I. Taufik dan S.A. Pamungkas

2	@Reiza_Patters @temanAhok Kalo gue bilang Ahok bego kalian mau apa? Gue punya KTP DKI hak gue buat nilai Gubernur bego dan gagal.	reizapatters temanahok kalo gue bilang ahok bego kalian mau apa gue punya ktp dki hak gue buat nilai gubernur bego dan gagal
3	@RatnaSpaet @SiBonekaKayu @basuki_btp @Aktualco itulah tipu daya ahok sungguh licik dan munafik...	ratnaspamet sibonekakayu basukibtp aktualco itulah tipu daya ahok sungguh licik dan munafik
4	INILAH AHOK @basuki_btp MUSANG BERBULU DOMBA KAFIR BERBUSANA MUSLIM. pic.twitter.com/v8bdLXkZGC	inilah ahok basuki_btp musang berbulu domba kafir berbusana muslim pic.twitter.com/v8bdlxk zgc

**Tabel 6.** Proses *Filtering*

No	<i>Tweet</i>	<i>Tokenizing</i>
1	@Ronni_Hermawan_ @basuki_btp @kangdede78 Alhamdulillah warga Bekasi bisa ke Jakarta pp dgn bis TJ hanya dgn 3500. Terima kasih Ahok	alhamdulillah bekasi jakarta bis

Analisis Sentimen terhadap Tokoh Publik menggunakan Algoritma *Support Vector Machine*

		tj 3500 terima kasih
2	@Reiza_Patters @temanAhok Kalo gue bilang Ahok bego kalian mau apa? Gue punya KTP DKI hak gue buat nilai Gubernur bego dan gagal.	bilang bego ktp dki hak gubernur bego gagal
3	@RatnaSpaet @SiBonekaKayu @basuki_btp @Aktualco itulah tipu daya ahok sungguh licik dan munafik...	tipu daya sungguh licik munafik
4	INILAH AHOK @basuki_btp MUSANG BERBULU DOMBA KAFIR BERBUSANA MUSLIM. pic.twitter.com/v8bdLXkZGC	inilah musang berbulu domba kafir berbusana muslim

**Tabel 7.** Proses *Stemming*

No	<i>Tweet</i>	<i>Tokenizing</i>
1	@Ronni_Hermawan_ @basuki_btp @kangdede78 Alhamdulillah warga Bekasi bisa ke Jakarta pp dgn bis TJ hanya dgn 3500. Terima kasih Ahok	alhamdulillah bekasi jakarta bis tj 3500 terima kasih
2	@Reiza_Patters @temanAhok Kalo gue bilang Ahok bego kalian mau apa? Gue punya KTP DKI hak gue buat nilai Gubernur bego dan gagal.	bilang bego ktp dki hak gubernur bego gagal
3	@RatnaSpaet @SiBonekaKayu @basuki_btp @Aktualco itulah tipu daya ahok sungguh licik dan munafik...	tipu daya sungguh licik munafik
4	INILAH AHOK @basuki_btp MUSANG BERBULU DOMBA KAFIR BERBUSANA MUSLIM.	ini musang bulu

## I. Taufik dan S.A. Pamungkas

	pic.twitter.com/v8bdLXkZGC	domba kafir busana muslim
--	----------------------------	------------------------------------

Pada proses ini dilakukan menghitung banyaknya *term* atau kata yang muncul pada *tweet* (tf), menghitung banyaknya *tweet* yang mengandung *term* tersebut (df), menghitung inverse dokumen *frequency* (idf), dan mengalikan tf dengan idf sebagai bobot dari *term* pada setiap *tweet*. Hasil pembobotan tertera pada Tabel 8.

**Tabel 8.** Hasil Pembobotan

<i>Term</i>	tf				df	n/df	idf	tf-idf			
	t <sub>1</sub>	t <sub>2</sub>	t <sub>3</sub>	t <sub>4</sub>				t <sub>1</sub>	t <sub>2</sub>	t <sub>3</sub>	t <sub>4</sub>
alhamdulillah	1	0	0	0	1	4	0,60	0,60	0,00	0,00	0,00
bekasi	1	0	0	0	1	4	0,60	0,60	0,00	0,00	0,00
jakarta	1	0	0	0	1	4	0,60	0,60	0,00	0,00	0,00
bis	1	0	0	0	1	4	0,60	0,60	0,00	0,00	0,00
tj	1	0	0	0	1	4	0,60	0,60	0,00	0,00	0,00
3500	1	0	0	0	1	4	0,60	0,60	0,00	0,00	0,00
terima	1	0	0	0	1	4	0,60	0,60	0,00	0,00	0,00
kasih	1	0	0	0	1	4	0,60	0,60	0,00	0,00	0,00
bilang	0	1	0	0	1	4	0,60	0,00	0,60	0,00	0,00
bego	0	2	0	0	1	4	0,60	0,00	1,20	0,00	0,00
ktp	0	1	0	0	1	4	0,60	0,00	0,60	0,00	0,00
dki	0	1	0	0	1	4	0,60	0,00	0,60	0,00	0,00
hak	0	1	0	0	1	4	0,60	0,00	0,60	0,00	0,00
gubernur	0	1	0	0	1	4	0,60	0,00	0,60	0,00	0,00
gagal	0	1	0	0	1	4	0,60	0,00	0,60	0,00	0,00
tipu	0	0	1	0	1	4	0,60	0,00	0,00	0,60	0,00
daya	0	0	1	0	1	4	0,60	0,00	0,00	0,60	0,00
sebenarnya	0	0	1	0	1	4	0,60	0,00	0,00	0,60	0,00
munafik	0	0	1	0	1	4	0,60	0,00	0,00	0,60	0,00
ini	0	0	0	1	1	4	0,60	0,00	0,00	0,00	0,60
musang	0	0	0	1	1	4	0,60	0,00	0,00	0,00	0,60
bulu	0	0	0	1	1	4	0,60	0,00	0,00	0,00	0,60
domba	0	0	0	1	1	4	0,60	0,00	0,00	0,00	0,60
kafir	0	0	0	1	1	4	0,60	0,00	0,00	0,00	0,60
busana	0	0	0	1	1	4	0,60	0,00	0,00	0,00	0,60
muslim	0	0	0	1	1	4	0,60	0,00	0,00	0,00	0,60

dengan  $t_i$  dokumen ke- $i$ , df *document frequency*, n banyaknya data, dan idf *Inverse document frequency*.





$$x_3 \cdot x_3' = 1.0125.$$

Diperoleh matriks  $K = \begin{bmatrix} 0.9516 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1.7325 \end{bmatrix}$ . Hasil ini disubsitusikan ke persamaan:

$$maks L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

dan diperoleh nilai parameter  $\alpha_1 = 1$ ,  $\alpha_2 = -0,56$ ,  $\alpha_3 = -0,44$  dan  $b = 1,74$ .

Selanjutnya dilakukan pengujian analisis sentimen dengan data *training* berjumlah 420 dan data *testing* berjumlah 210 sehingga, total semua data berjumlah 630 data. Tabel 10 merupakan hasil pengklasifikasian menggunakan SVM.

**Tabel 10.** Hasil Pengklasifikasian SVM

No	Label Data	Nilai Probabilitas
1	Positif	0,52
2	Positif	0,93
3	Positif	0,55
4	Positif	0,54
5	Positif	0,76
⋮	⋮	⋮
210	Positif	0,75

Hasil analisa sentimen menggunakan Algoritma SVM dengan penggunaan data sebanyak 630 yang terbagi atas 420 *data training* dan 210 data *testing* dievaluasi dengan dihitung nilai akurasi, presisi dan *recallnya*. Tabel 11 menjelaskan tentang matriks konfusi hasil klasifikasi sentimen.

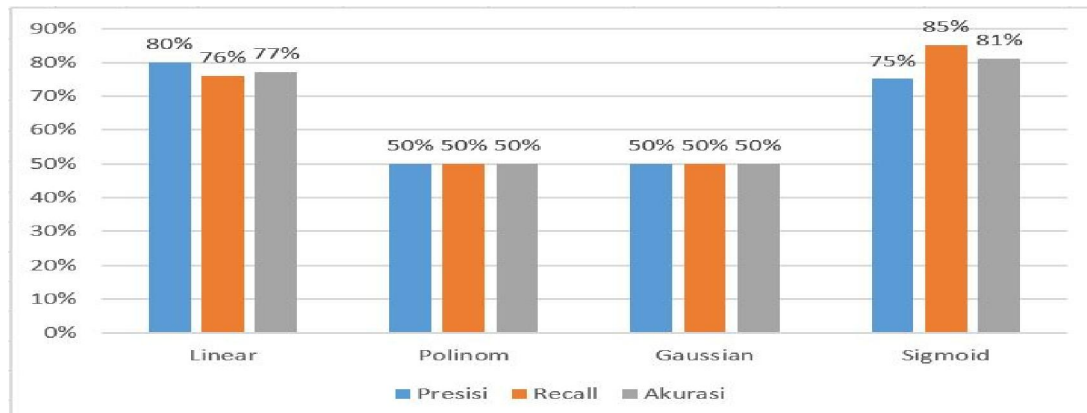
**Tabel 11.** Hasil Matriks Konfusi

		Kelas Hasil Prediksi		Total
		Negatif	Positif	
Kelas Asli	Negatif	75	27	105
	Positif	21	54	105

Hasil perhitungan SVM juga dapat dipengaruhi oleh pemilihan parameter. Parameter di sini bisa berarti nilai C atau nilai kernel. Hasil perbandingan terhadap parameter kernel. *Kernel linear*, *Kernel polinomial* dan juga *Kernel Radial* atau disebut juga *Kernel gaussian* dan *Kernel sigmoid* tertera pada Tabel 12.

**Tabel 12.** Perbandingan Evaluasi *Kernel linear*, *Polinom*, *Gaussian* dan *Sigmoid*.

	<i>Linear</i>	<i>Polinom</i>	<i>Gaussian</i>	<i>Sigmoid</i>
<i>Presisi</i>	80%	50%	50%	75%
<i>Recall</i>	76%	50%	50%	85%
<i>Akurasi</i>	77%	50%	50%	81%



Gambar 3. Perbandingan Kernel SVM

## KESIMPULAN

Penelitian ini telah berhasil mengimplementasikan Algoritma SVM untuk analisa sentimen pada data *tweet* tentang tokoh publik dengan *keyword* “Ahok” dan “@teman\_ahok”. Hal ini diperlihatkan dengan hasil akurasi menggunakan data *tweet* sebanyak 630 data, dimana 420 data adalah data *training* dan 210 adalah data testing. Hasil menunjukkan bahwa *Kernel Linear* memiliki tingkat presisi paling baik sekitar 80% dilanjutkan dengan *Kernel Sigmoid* sekitar 75%. Untuk *Kernel Gaussian* dan Polinom nilai akurasinya sekitar 50%. *Kernel sigmoid* memiliki tingkat *recall* paling baik sekitar 85% dilanjutkan dengan *Kernel Linear* sekitar 76%. Untuk *Kernel Gaussian* dan Polinom nilai akurasinya sekitar 50%. *Kernel Sigmoid* memiliki tingkat akurasi paling baik sekitar 81% dilanjutkan dengan *Kernel Linear* sekitar 77%. Untuk *Kernel Gaussian* dan Polinom nilai akurasinya sekitar 50%.

## REFERENSI

- [1] Agarwal, A., B. Xie., I. Vovsha, O. Rambow and R. Passonneau, 2011, Sentiment Analysis of Twitter Data, Language in Social Media, Oregon: *Association for Computational Linguistics*, pp. 30-38.
- [2] Aliandu, P., 2012, *Analisis Sentimen Tweet Berbahasa Indonesia di Twitter*, Thesis Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Gajah Mada. Yogyakarta.
- [3] Andrew, W. M., 2011, *Support Vector Machine*, Carniege: Carniege Melon University.
- [4] Berry, M W. and J. Kogan, 2010, *Text Mining Applications and Theory*. Sussex: Wiley.
- [5] Feldman, R., and J. Sanger, 2007, *The Text Mining Handbook : Advance Approaches in Analyzing Unstructured Data*, New York: Cambride University Press.
- [6] Liu, B., 2012, *Sentiment Analysis and Opinion Mining*. San Rafael: Morgan & Claypool Publisher.
- [7] Maning, D.C., P. Raghavan, and H. Schutze, 2008, *Introduction to Information Retrieval*, Cambridge University Press.
- [8] Pang, B., L. Lee, and S. Vaithyanathan, 2002, Thumbs up? Sentiment Classification using Machine Learning Techniques, *Proceedings of EMNLP*, pp. 79-86.
- [9] Romelta, E., 2012, Opinion Mining di Twitter untuk Customer Feedback Smartphone dengan Pembelajaran Mesin, *Jurnal Sarjana Institut Teknologi Bandung Bidang Teknik Elektro dan Informatika*, Vol. 1, No 2.
- [10] Vidya, N.A., M.I. Fanany, dan I. Budi, 2015, Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers, *Procedia Computer Science*, pp. 519-526.