SENTIMEN ANALYSIS PENGGUNA *TWITTER* TERHADAP PASANGAN CALON GUBERNUR JAWA TIMUR TAHUN 2018 MENGGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE*

[1] Apriyono Thoyyib (1410651111)
[2] Triawan Adi Cahyanto, M.Kom
Program Studi Teknik Informatika,
Universitas Muhammadiyah Jember
Email : apriyono.t@gmail.com

ABSTRACT

The Governor and Deputy Governor election of East Java in 2018 is widely discussed in the real world and cyberspace, especially on Twitter social media. Everyone are free to have an opinion about a candidate for Governor, so there are many positive opinions and negative opinions. Social media is also used as a campaign which he thinks is efficient. This research is expected to make people understand the sentiments that contain positive or negative. In the first stage, preprocessing will be done first, namely using case folding, stopword removal, tokenizing and stemming. Then for the classification process using Support Vector Machine (SVM). The data is 200 used is a tweet in Indonesian with the keyword khofifahemil and gusipulmbakputi. The results of this study are sentiment analysis of the candidates for the East Java Governor in 2018. So as to obtain an accuracy of 95% with positive sentiment as many as 19 data and negative sentiments as much as 1 data for the Khofifah and Emil pairs, while 90% with positive sentiments as much as 18 data and negative sentiments as much as 2 data for couples Gus Ipul and Puti.

***Keywords** : sentiment analysis, east java governor candidate 2018 , preprocessing, support vector machine.*

ABSTRAK

Pemilihan Calon Gubernur dan Wakil Gubernur Jawa Timur tahun 2018 ramai diperbincangkan di dunia nyata dan dunia maya, khususnya di media sosial Twitter. Semua orang bebas berpendapat atau beropini tentang calon Gubernur, sehingga banyak opini yang positif dan opini negatif. Media sosial juga dijadikan sebagai mendia untuk berkampanye yang menurutnya efisien. Penelitian ini diharapkan dapat menjadikan masyarakat paham akan sentimen yang mengandung positif atau negatif. Pada tahap pertama akan dilakukan *preprocessing* terlebih dahulu yaitu menggunakan *case folding, stopword removal, tokenizing dan stemming*. Selanjutnya untuk proses klasifikasinya menggunakan *Support Vector Machine (SVM)*. Data yang digunakan adalah *tweet* dalam bahasa Indonesia dengan kata kunci khofifahemil dan gusipulmbakputi, dengan jumlah data set sebanyak 200 *tweet*. Hasil dari penelitian ini adalah analisis sentimen terhadap calon pasangan Gubernur Jawa Timur tahun 2018. Sehingga diperoleh akurasi sebesar 95 % dengan sentimen positif sebanyak 19 data dan sentimen negatif sebanyak 1 data untuk pasangan Khofifah dan Emil, sedangkan 90 % dengan sentimen positif sebanyak 18 data dan sentimen negatif sebanyak 2 data untuk pasangan Gus Ipul dan Puti.

Kata kunci *: analisis sentimen, calon gubernur jawa timur tahun 2018, preprocessing, support vector machine.*

1. Pendahuluan

Pemilihan Gubernur akan dilaksanakan secara serentak pada bulan Juni tahun 2018. Pilkada secara langsung diberlakukan sejak adanya Undang-Undang nomor 32 tahun 2004. Ditegaskan dalam UU ini, pemilu dilaksanakan berdasarkan asas langsung, umum, bebas, rahasia, jujur, dan adil. Menjelang pemilihan gubernur pada tahun 2018 ini (Setkab.go.id), tentunya para calon saling mengumbar janji secara langsung atau melalui media sosial untuk berkampanye demi peningkatan suara mereka.

Media sosial sekarang ini menjadi perangkat komunikasi yang sangat poluler di kalangan pengguna internet. Contoh dari media sosial yaitu Facebook, Twitter, Path, Instagram, Google+, Tumblr dan lain sebagainya (Evadollzz, 2014). Media sosial yang populer saat ini yaitu Twitter, tidak hanya digunakan sebagai sarana pertemanan, mencari teman, tapi sudah banyak digunakan sebagai sarana promo dagangan bahkan kampanye calon Gubernur atau Presiden.

Tim sukses dalam memenangkan sebuah pasangan calon gubernur, dengan melakukan kampanye baik langsung atau melalui sosial media. Salah satu media sosial yang banyak digunakan adalah Twitter. Dua pasangan calon tersebut juga aktif di sosial media khususnya Twitter, pasangan pertama adalah Dra.Khofifah Indar Parawansa dengan nama akun @khofifahIP dan pasangan kedua adalah Drs.Syaifullah Yusuf (Gus Ipul) dengan nama akun @GusIpul_ID.

Beberapa lembaga survei telah mengeluarkan hasilnya diantaranya Charta Politika yang menyatakan dalam rilisnya jika paslon Pilkada Jatim nomor urut dua unggul dari pesaingnya, sebanyak 44,8 persen dikantongi paslon nomor 2 dan 38,1 persen dikantongi paslon nomor 1. Sementara lembaga survei *Poltracking* menyampaikan hasil survei elektabilitas paslon nomor 1 lebih unggul sebesar 42,4 persen sedangkan paslon nomor 2 mendapatkan 35,8 persen (Marlinda oktavia erwanti, news.detik.com).

Kedua lembaga survei saling memenangkan paslon dengan *poolling* yang berbeda. Keberpihakan kepada salah satu paslon ini membuat masyarakat terpengaruh untuk menentukan pilihannya.

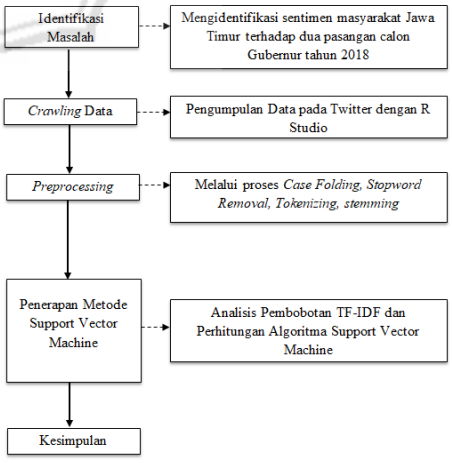
Text mining adalah proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari sejumlah besar sumber data tak terstruktur. Penambangaan teks memiliki tujuan dan menggunakan proses yang sama dengan penambangan data, namun memiliki masukan yang berbeda. Penambangan teks dapat dianggap sebagai

proses dua tahap yang diawali dengan penerapan stuktur terhdap sumber data teks dan dilanjutkan ekstraksi informasi dan pengetahuan yang relevan, sementara *sentiment analysis* merupakan proses memahami, mengestrak dan mengolah data tekstual secara otomatis untuk medapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. *Sentiment analysis* dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, cenderung positif atau negatif (Marfian, 2015). Pada penelitian ini analisis sentimen dilakukan untuk melihat sebuah opini sesorang yang ditujukan kepada calon Gubernur Jawa Timur 2018, opini itu bisa dimasukkan kategori opini positif atau negatif. Besarnya sentimen yang ditujukan kepada calon Gubernur Jawa Timur 2018 bisa dijadikan sebuah parameter kemenangan atau kekalahan seseorang calon.

Penelitian ini memanfaatkan Twitter dengan menganalisis *tweet* berbahasa indonesia yang membicarakan tokoh publik menjelang pemilihan gubernur Jawa Timur 2018 di Indonesia. Dua pasangan calon Gubernur tersebut akan dianalisis untuk menentukan kecenderungan masyarakat terhadap sebuah pasangan calon. Analisis dilakukan berdasarkan akseptabilitas setiap pasangan calon gubernur, akseptabilitas adalah sikap penerimaan masyarakat terhadap seorang tokoh (Santosa, 2014). Terdapat banyak metode klasifikasi diantaranya *K-Means, Apriori, Naive Bayes, Support Vector Machine* dan lainnya. Pada penelitian ini digunakan metode *Suppport Vector Machine* karena pada penelitian yang telah dilakukan memiliki tingkat keakuratan tinggi dan kelebihan metode ini yaitu dapat menentukan *hyperlane* atau bidang pemisah dengan memilih bidang dengan optimal *margin* maka generalisasi dapat terjaga dengan sendirinya, tingkat generalisasi pada *Support Vector Machine* tidak dipengaruhi oleh jumlah data latih dengan menentukan parameter *soft margin*, *noise* dapat dikontrol sehingga proses pelatihan ketat.

2. Metode penelitian

Dalam pengerjaan Tugas Akhir ini diperlukan langkah-langkah kegiatan penelitian untuk mendapatkan hasil yang maksimal. Adapun langkah-langkah penelitian untuk analisis sentimen pada Twitter pada elektabilitas pasangan calon Gubernur Jawa Timur tahun 2018 dengan menggunakan metode Support Vector Machine adalah sebagai berikut :



2.1 Identifikasi Masalah

Identifikasi masalah adalah satu proses penelitian yang bisa dikatakan langkah

penting pada proses identifikasi. Dalam masalah ini akan fokus terhadap sentimen masyarakat Jawa Timur pada Twitter terhadap dua pasangan calon Gubernur tahun 2018. Tujuannya untuk mengklasifikasikan sentimen seluruh masyarakat khususnya pada opini pada Twitter.

2.2 Pembuatan Kamus Data

Tahap ini adalah menyusun kamus data positif dan negatif dari analisis data *tweet* yang sudah divalidasi oleh Ahli.

Tabel 3.1 Potongan Hasil Analisis Dan Pengumpulan Kamus Sentimen Positif

Berjanji	Membangun	Kemenangan	kebaikan
Kedamaian	Potensi	Lancar	berkomitmen
Pejuang	Aqidah	Akhlak	istimewa
mengunjungi	terimakasih	Alhamdulillah	meningkatkan
Cinta	Kebaikan	Keikhlasan	mengunjungi
Konsolidasi	Deklarasi	lebih maju	lebih baik
Bangkit	Berkarya	Insyaallah	sejahtera
Yakin	lebih makmur	Berpengalaman	idaman
Idola	Ziarah	Teladan	sejahtera
Amanah	Jujur	Berjuang	Bersih
Cerdas	do'a	Bukti	kemajuan
Pintar	Cerah	gotong royong	kharisma
menghargai	Kelancaran		

Tabel 3.2 Potongan Hasil Analisis Dan Pengumpulan Kamus Sentimen Negatif

Jelek	kurang terawat	tidak ada	panas
Rusak	Mengganggu	Masalah	kemiskinan
pengangguran	sebelah mata	Semerawut	tercemar
Kotoran	Mangkrak	Bingung	Takut
Miris	tidak merata	Menjerit	kurang layak
Pemborosan	Memanfaatkan	Masalah	dirusak
Hoax	Menyimpang	Keresahan	pemborosan
Bukan	Mengkritik	Kebencian	Ragu
tidak kondusif	Jangan		

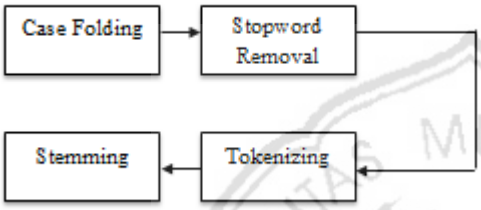
2.3 Crawling data

Tahap crawling data adalah tahap dikumpulkan semua data hasil opini masyarakat terhadap dua pasangan calon Gubernur Jawa Timur tahun 2018 dengan kata kunci hashtag : gusipulmbakputi, khofifahemil. Pertama utnuk melakukan *crawling* data terlebih dahulu harus mendaftarkan aplikasi yang dibuat ke Twitter untuk mendapatkan user *credential* yang nantinya digunakan dalam proses *crawling* data dari Twitter. Proses ini bertujuan untuk mendapatkan *authentication* dari Twitter terhadap akses data yang dimiliki oleh Twitter (Negara, 2016). Kedua mengumpulkan data Twitter menggunakan R Studio dengan melakukan pemanggilan hashtag tersebut. Pengumpulan dari penelitian ini yaitu data yang diunduh berupa *user* dan *tweet*

beserta atribut-atributnya, lalu disimpan dalam bentuk file csv.

2.4 Preprocessing

Agar proses klasifikasi sentimen tersebut berjalan dengan baik, maka diperlukan proses prepocessing. Dalam proses ini akan dilakukan pembersihan *noise* seperti *link*, "RT", "@" dan *incomplete* yaitu tidak lengkapnya suatu atribut serta *incosistent* yaitu nilai yang saling bertentangan.



2.5 *Case folding*

Case folding merupakan tahapan yang mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter atau pembatas (Larasati, 2015). Seperti contoh dibawah ini :

Input :

@jember Sudut yang berbeda dari Spot Foto Baru di Pantai Papuma

@jember Kabar menggembirakan bagi masyarakat Jember ada Spot baru

RT @jember Raja Ampat-nya Jember

Output :

sudut yang berbeda dari spot foto baru di pantai papuma

kabar menggembirakan bagi masyarakat jember ada spot baru

raja ampat-nya jember

2.6 Stopword Removal

Stopword Removal merupakan proses *cleansing* yang masih tersisa jadinya perlu dilakukan *stopword* untuk menghilangkannya. Dalam *stopword* biasanya berupa kata ganti orang, kata ganti penghubung dan lain sebagainya. Contoh *stopword* untuk Bahasa Inggris diantaranya "of", "the". Sedangkan untuk bahasa Indonesia diantaranya "yang", "di", "ke" (Larasati, 2015). Seperti contoh dibawah ini :

ada	Bagi	di	jadi	maka	pasti	ucap
adalah	Bagian	entah	jangan	makin	punya	ujar
adanya	Bahkan	gunakan	juga	mana	rasa	umum
adapun	Bahwa	hampir	kembali	mulai	saat	usai
agak	Cara	hanya	kapan	namun	saling	wahai
agaknya	Cukup	hingga	ke	nanti	sangat	walau
agar	Dan	ikut	lagi	oleh	sebab	yaitu

Input :

sudut **yang** berbeda **dari** spot foto baru **di** pantai papuma

kabar menggembirakan **bagi** masyarakat jember ada spot baru

raja ampat-nya jember

Output :

sudut berbeda spot foto baru pantai papuma

kabar menggembirakan masyarakat jember ada spot baru

raja ampat-nya jember

2.7 Tokenizing

Tokenizing merupakan proses pemotongan string input berdasarkan tiap kata penyusunanya (Larasati, 2015). Selain itu, spasi digunakan untuk memisahkan antar kata tersebut. Seperti contoh dibawah ini :

Input :

sudut berbeda spot foto baru pantai papuma

kabar menggembirakan masyarakat jember ada spot baru

raja ampat-nya jember

Output :

sudut	berbeda	spot	foto	baru	pantai	papuma
kabar	menggembirakan	masyarakat	jember	ada		
spot	baru					
raja	ampat-nya	jember				

2.8 Stemming

Stemming merupakan bagian yang tidak terpisahkan

dalam *Information Retrieval* (*IR*). Tidak banyak algoritma yang dikhususkan untuk *stemming*bahasa Indonesia dengan berbagai keterbatasan didalamnya. Algoritma Porter salah satunya, Algoritma ini membutuhkan waktu yang lebih singkat dibandingkan dengan *stemming* menggunaka n Algoritma Nazief & Adriani, namun proses *stemming* menggunakan Algoritma Porter memiliki prosentase keakuratan (presisi) lebih kecil dibandingkan dengan stemming menggunakan Algoritma Nazief & Adriani. Algoritma Nazief & Adriani sebagai algoritma *stemming* untuk teks berbahasa Indonesia yang memiliki kemampuan prosentase keakuratan (presisi) lebih baik dari algoritma lainnya. Algoritma ini sangat dibutuhkan dan menentukan dalam proses IR dalam dokumen Indonesia (Teguh, 2017).

Stemming adalah salah satu cara yang digunakan untuk meningkatkan performa IR dengan cara mentransformasi kata-kata dalam sebuah dokumen teks ke bentuk kata dasarnya. Algoritma *stemming* untuk bahasa yang satu berbeda dengan algoritma stemming untuk bahasa lainnya. Sebagai contoh bahasa Inggris memiliki morfologi yang berbeda dengan bahasa Indonesia sehingga algoritma *stemming* untuk kedua bahasa tersebut juga berbeda. Proses *stemming* pada teks berbahasa Indonesia lebih rumit/kompleks karena terdapat variasi imbuhan yang harus dibuang untuk mendapatkan *root word*(kata dasar) dari sebuah kata. Gambaran proses dari tahap *stemming* dapat dilihat pada Gambar 3.6.

Langkah-langkah algoritma Nazief dan Adriani adalah sebagai berikut:

1) Kata yang belum di-*stemming* dicari pada kamus, jika ditemukan, kata tersebut dianggap sebagai kata dasar yang benar dan algoritma dihentikan.
2) Hilangkan *Inflectional suffixes*, yaitu dengan menghilangkan *particle*(“-lah”, ”-kah”, “tah” atau “-pun”), kemudian hilangkan *inflectional possessive pronoun suffixes* (“-ku”,“-mu” atau ”-nya”). Cek kata di dalam kamus kata dasar, jika ditemukan, algortima dihentikan, jika tidak lanjut ke langkah 3. 3) Hapus *Derivational Suffix* (“-i” atau ”-an”,”). Jika kata ditemukan dalam kamus kata dasar, maka algoritma berhenti. Jika tidak, maka lanjut ke langkah 3a:
    a. Jika akhiran “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “-k”, maka “-k” juga dihapus. Jika kata tersebut ditemukan dalam kamus maka algoritma berhenti. Jika tidak ditemukan maka lakukan langkah 3b.
    b. Akhiran yang dihapus (“-i”, “- an” atau “-kan”)

dikembalikan, lanjut ke langkah 4.

c. Hapus *Derivational Prefix* (“be-”,”di-”,”ke-”,”me-”,”pe-“,”se-” dan “te-“). Jika kata yang didapat ditemukan didalam database kata dasar, maka proses dihentikan, jika tidak, maka lakukan *recoding*.

Tahapan ini dihentikan jika memenuhi beberapa kondisi berikut:

a. Terdapat kombinasi awalan dan akhiran yang tidak diijinkan
b. Awalan yang dideteksi sama dengan awalan yang dihilangkan sebelumnya.
c. Tiga awalan telah dihilangkan

3) Jika semua langkah telah dilakukan tetapi kata dasar tersebut tidak ditemukan pada kamus, maka algoritma ini mengembalikan kata yang asli sebelum dilakukan *stemming*.

Tipe awalan ditentukan melalui langkah-langkah berikut:

1. Jika awalannya adalah: “di-”, “ke-”, atau “se-” maka tipe awalannya secara berturut-turut adalah “di-”, “ke-”, atau “se-”.
2. Jika awalannya adalah “te-”, “me-”, “be-”, atau “pe-” maka dibutuhkan sebuah proses tambahan untuk menentukan tipe awalannya.
3. Jika dua karakter pertama bukan “di-”, “ke-”, “se-”, “te-”, “be-”, “me-”, atau “pe-” maka berhenti.
4. Jika tipe awalan adalah “none” maka berhenti. Jika tipe awalan adalah bukan “none” maka awalan dapat dilihat pada Tabel 2. Hapus awalan jika ditemukan.

Tabel 3.3 Kombinasi awalan dan akhiran yang tidak diijinkn

Awalan	Akhiran yang tidak diizinkan
be-	-i
di-	-an
ke-	-i, -kan
me-	-an
se-	-i, -kan

Tabel 3.4 Cara menentukan tipe awalan untuk awalan “ter-“

Following Characters				Tipe Awalan
Set 1	Set 2	Set 3	Set 4	
“-r-“	“-r-“	–	–	None
“-r-“		–	–	ter-luluh
“-r-“	not (vowel or “-r-”)	“-er-“	vowel	Ter
“-r-“	not (vowel or “-r-”)	“-er-“	not vowel	ter-
“-r-“	not (vowel or “-r-”)	not “-er-“	–	Ter

not (vowel or "-r-")	"-er-"	Vowel	–	None
not (vowel or "-r-")	"-er-"	not vowel	–	Te

Tabel 3.5 Jenis awalan berdasarkan tipe awalannya

Tipe Awalan	Awalan yang harus dihapus
di-	di-
ke-	ke-
se-	se-
te-	te-
ter-	ter-
ter-luluh	Ter

Untuk mengatasi keterbatasan pada algoritma di atas, maka ditambahkan aturan-aturan dibawah ini:

1. Aturan untuk reduplikasi.
   a. Jika kedua kata yang dihubungkan oleh kata penghubung adalah kata yang sama maka *root word* adalah bentuk tunggalnya, contoh : "buku-buku" *root word*-nya adalah "buku".
   b. Kata lain, misalnya "bolak-balik", "berbalas-balasan, dan "seolah-olah". Untuk mendapatkan *root word*-nya, kedua kata diartikan secara terpisah. Jika keduanya memiliki *root word* yang sama maka diubah menjadi bentuk tunggal, contoh: kata "berbalas-balasan", "berbalas" dan "balasan" memiliki *root word* yang sama yaitu "balas", maka *root word* "berbalas-balasan" adalah "balas". Sebaliknya, pada kata "bolak-balik", "bolak" dan "balik" memiliki *root word* yang berbeda, maka *root word*-nya adalah "bolak-balik".
2. Tambahan bentuk awalan dan akhiran serta aturannya.
   a. Untuk tipe awalan "mem-", kata yang diawali dengan awalan "memp-" memiliki tipe awalan "mem-".
   b. Tipe awalan "meng-", kata yang diawali dengan awalan "mengk-" memiliki tipe awalan "meng-".

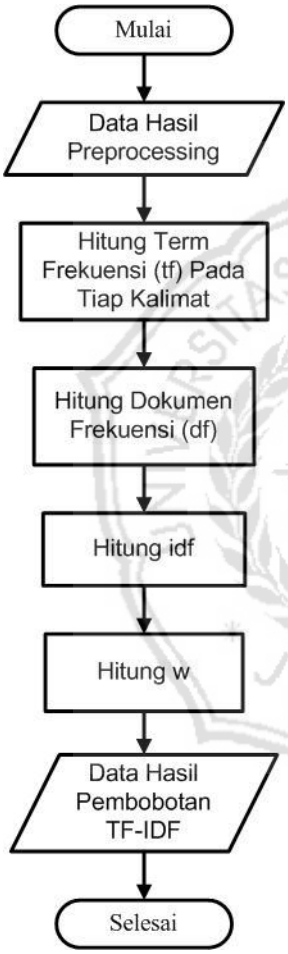
Contoh *stemming* dibawah ini sebagai berikut :

1. Awalan meng- dan berakhiran –kan
   Contoh meng-gembira-kan
2. Awalan ber-
   Contoh ber-beda

3. Analisis Pembobotan TF-IDF

Pembobotan kata digunakan untuk mendapatkan suatu topik atau *keyword* dari kumpulan teks. Salah satu metode pembobotan adalah *Term Frequency-Inverse Document Frequency* (TF-IDF). Pada penelitian ini pembobotan diperoleh dari

jumlah kemunculan *term* dalam satu dokumen *term frequency* (tf) dan sebuah kata di dalam kumpulan dokumen atau jumlah kemunculan *term* dalam koleksi dokumen *inverse document frequency* (idf) (Citrawatii, 2011). Adapun *flowchart* dari proses pembobotan tf-idf dapat dilihat pada Gambar 3.11 berikut :



Gambar 3.1 *Flowchart* Pembobotan TF-IDF

Berikut adalah data latih yang sudah melewati tahapan *preprocessing* dengan contoh 3 dokumen dapat dilihat pada Tabel 3.7.

Tabel 3.6 Tabel Pernyataan Latih

Pernyataan	**Kata**
P1	calon butuh masyarakat jatim gusipul mbakputi karakter anti korupsi islam ayom
P2	keringat himpun dukung jokowi jatim pilpers2014 khofifah wajar kalau pilih jokowi pro khofifah
P3	tak layak calon klaim dukung jokowi

Contoh disini akan diterapkan terhadap kata "calon" pada pernyataan pertama (P1).

3.1 Menghitung *Term Frequency* (TF)

Hasil data yang sudah di lakukan proses *preprocessing* dihitung berdasarkan kemunculannya contoh disini yaitu kata "calon". Dokumen pada penelitian ini adalah P1,P2 dan P3.

Tabel 3.7 Tabel *Frequency* (TF)

Kata	***Term Frequency (tf)***		
	P1	**P2**	**P3**
calon	1	0	0

Dari tabel 3 menjelaskan bahawa kata calon pada pernyataan pertama (P1) muncul satu kali, sedangkan (P2) dan (P3) kata "calon" tidak muncul sama sekali (0).

3.2 Menghitung Document Frequency (df)

Nilai *document frequency* (df) didapatkan dari jumlah dokumen

yang terdapat kata "calon" yaitu 1 pada dokumen (P1) saja.

3.3 Menghitung Inverse Document Frequency (idf)

Setelah mendapatkan nilai *df*, maka dilakukan perhitungan *idf* dengan persamaan (2.3).

Diambil contoh pada kata "calon". Didapatkan banyak dokumen (N) =3, dan df=1. Maka dihitung seperti berikut, hasilnya sebagai berikut :

$$\text{IDF} = \log\left(\frac{\text{N}}{\text{df}}\right)$$

$$= \log\left(\frac{3}{1}\right)$$

$$= \log(3) = 0.477$$

N merupakan jumlah dokumen, pada penelitian ini terdapat 3 dokumen yaitu (P1), (P2) dan (P3). Kemudian nilai df pada kata "calon" didapatkan 0.477.

3.4 Menghitung W

Menghitung bobot (w) kata "calon" pada setiap pertanyaan dapat dihitung menggunakan persamaan sebagai berikut (2.4).

Maka hasilnya sebagai berikut :

$$w_{p1} = tf_{p1} \text{ X } idf = 1 \text{ X } 0.477$$

$$= 0.477$$

$$w_{p2} = tf_{p2} \text{ X } idf = 0 \text{ X } 0.477$$

$$= 0$$

$$w_{p3} = tf_{p3} \text{ X } idf = 0 \text{ X } 0.477$$

$$= 0$$

Dari perhtitungan (w) yang telah dilakukan, diperoleh nilai (w) untuk P1 adalah 0.477 dan (P2) dan (P3) 0.

Nilai (P1) adalah 1 di dapatkan dari kata "calon", sementara di (P2),(P3) tidak ada kata yang serupa maka di berikan nilai 0. Nilai df didapat dari jumlah kata sama yang terdapat dalam satu dokumen. Nilai N/df didapatkan dari jumlah dokumen/jumlah kata yang muncul dalam satu dokumen tersebut sehingga didapatkan nilai idf sebesar 0.477. Sementara nilai w didapatkan dari kemunculan kata dalam dokumen dikalikan dengan idf seperti dalam tabel dibawah ini. Berikut hasil perhitungan pembobotan TF-IDF keseluruhan kata dan latih hasil *preprocessing* dapat dilihat pada Tabel 3.9.

Tabel 3.8 Tabel Perhitungan Pembobotan TF-IDF Latih

Kata	Tf			df	N/df	idf	W		
	P1	P2	P3				P1	P2	P3
Calon	1	0	0	1	3/1	0.477	0.477	0	0
Butuh	1	0	0	1	3/1	0.477	0.477	0	0
masyarakat	1	0	0	1	3/1	0.477	0.477	0	0
Jatim	1	1	0	2	3/2	0.176	0.176	0.176	0
Gusipul	1	0	0	1	3/1	0.477	0.477	0	0
Mbakputi	1	0	0	1	3/1	0.477	0.477	0	0
Karakter	1	0	0	1	3/1	0.477	0.477	0	0
Anti	1	0	0	1	3/1	0.477	0.477	0	0
Korupsi	1	0	0	1	3/1	0.477	0.477	0	0
Islam	1	0	0	1	3/1	0.477	0.477	0	0
Ayom	1	0	0	1	3/1	0.477	0.477	0	0
Keringat	0	1	0	1	3/1	0.477	0	0.477	0
Himpun	0	1	0	1	3/1	0.477	0	0.477	0
Jokowi	0	2	1	3	3/3	0	0	0	0
pilpers2014	0	1	0	1	3/1	0.477	0	0.477	0
Khofifah	0	2	0	2	3/2	0.176	0	0.352	0
Wajar	0	1	0	1	3/1	0.477	0	0.477	0
Kalau	0	1	0	1	3/1	0.477	0	0.477	0
Pilih	0	1	0	1	3/1	0.477	0	0.477	0
Pro	0	1	0	1	3/1	0.477	0	0.477	0
tak	0	0	1	1	3/1	0.477	0	0	0.477
Layak	0	0	1	1	3/1	0.477	0	0	0.477
Calon	0	0	1	1	3/1	0.477	0	0	0.477
Klaim	0	0	1	1	3/1	0.477	0	0	0.477
Dukung	0	0	1	1	3/1	0.477	0	0	0.477

3.5 Perhitungan Algoritma Support Vector Machine

Tabel 3.9 Tabel Perhitungan Support Vector Machine

Term	TF	Kof-emil	kof-emil	Gus-Puti	Gus-Puti	DF	IDF
	Q	(+)	(-)	(+)	(-)		
Jawa	1	1	0	1	1	4	1,097
Timur	1	1	0	1	1	4	1,097
Cukup	1	0	0	1	0	2	1,398
Jadi	1	1	1	1	1	3	1,222
Beri	1	1	0	1	0	3	1,222
Kami	1	1	0	1	1	4	1,097
Selesai	1	0	0	0	0	1	1,699

Term disini adalah pecahan setiap kata di salah satu dokumen. TF adalah kemunculan kata di setiap dokumennya. Kemudian dihitung

setiap kata positif dan kata negatifnya antara Khofifah - Emil dan Gus Ipul – Puti, setiap kata diberikan nilai 1 sesuai dengan rumus yang sudah ditulis sebelumnya. DF adalah kemunculan kata yang sama diseluruh dokumen. IDF adalah jumlah TF, positif dan negatif dibagi DF kemudian di Log10 maka hasilnya untuk kata jawa adalah sebesar 1,097.

Tabel 3.10Tabel Kata Positif dan Kata Negatif

Idf	Kof-emil (+)	kof-emil (-)	Gus-Puti (+)	Gus-Puti (-)
1,097	1,097	0	1,097	1,097
1,097	1,097	0	1,097	1,097
1,398	0	0	1,398	0
1,222	1,222	1,222	1,222	1,222
1,222	1,222	0	1,222	0
1,097	1,097	0	1,097	1,097
1,699	0	0	0	0
Jumlah	5,735	1,222	7,133	4,513

Kemudian nilai IDF yang di dapatkan langsung dimasukkan sesuai kata positif dan kata negatif yang tadi sudah diberikan nilai. Selanjutnya setelah selesai maka dijumlahkan antara kata positif dan negatif Khofifah – Emil, Gus Ipul – Puti, yang diambil disini adalah angka terkecil pada proses ini kata tersebut masuk kedalam kata negatif.

4. Analisa

Pengujian metode *Support Vector Machine* dilakukan untuk menentukan sentimen masyarakat terhadap calon pasangan Gubernur Jawa Timur Tahun 2018, maka untuk mengehatahui akurasinya dilakukan perhitungan sebagai berikut :

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} x\ 100\%$$

Keterangan :

TP = (*True Positive*), yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.

TN = (*True Negative*), yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.

FN = (*False Negative*), yaitu jumlah data negatif namun terklasifikasi salah oleh sisten.

FP = (*False Positive*), yaitu jumlah data positif namun terklasifikasi salah oleh sistem.

Sehingga perhitungannya akurasinya dapat di gambarkan sebagai berikut untuk masing-masing pasangan calon Gubernur :

1. Pasangan Calon Gubernur Khofifah dan Emil

Confusion Matrix :

		Kelas Hasil Aplikasi	
		P	N
Kelas Sebenarnya	P	TP = 19	FN = 0
	N	FP = 1	TN = 0

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} x\ 100\%$$

$$Akurasi = \frac{19 + 0}{19 + 1 + 0 + 0} x\ 100\% = 95\ \%$$

2. Pasangan Calon Gubernur Gus Ipul dan Mbak Puti

Confusion Matrix :

	Kelas Hasil Aplikasi	
	P	N

Kelas Sebenarnya	P	TP = 16	FN = 0
	N	FP = 2	TN = 2

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} x\ 100\%$$

$$Akurasi = \frac{16 + 2}{16 + 2 + 2 + 0} x\ 100\%$$

$$= 90\ \%$$

Dengan demikian dapat disimpulkan bahwa hasil akurasi sebesar 95% dengan sentimen positif sebanyak 19 data dan sentimen negatif sebanyak 1 data untuk pasangan Khofifah dan Emil, sedangkan 90% dengan sentimen positif sebanyak 18 data dan sentimen negatif sebanyak 2 data untuk pasangan Gus Ipul dan Puti.

5. KESIMPULAN DAN SARAN

Dari hasil penelitian terhadap sentimen masyarakat pada calon Gubernur Jawa Timur Tahun 2018 dapat diambil beberapa kesimpulan dan saran sebagai berikut :

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilaksanakan, maka dapat disimpulkan sebagai berikut :

1. Algoritma *Support Vector Machine* dapat digunakan dalam mengklasifikasikan data *tweet* untuk analisis sentimen pada *twitter* terhadap calon pasangan Gubernur Jawa Timur tahun 2018.
2. Berdasarkan pengujian yang dilakukan dengan 160 data training dan 40 data uji masing – masing 20 data untuk setiap pasangan calon, dihasilkan akurasi akurasi sebesar 95% dengan sentimen positif sebanyak 19 data dan sentimen negatif sebanyak 1 data untuk pasangan Khofifah dan Emil, sedangkan 90% dengan sentimen positif sebanyak 18 data dan sentimen negatif sebanyak 2 data untuk pasangan Gus Ipul dan Puti.

5.2 Saran

Beberapa saran untuk pengembangan penelitian dimasa akan datang adalah sebagai berikut:

1. Dapat dikembangkan dengan metode yang baru untuk membandingkan agar mendapatkan nilai akurasi yang lebih baik.
2. Penambahan koleksi kamus pada database kata gaul atau singkatan, karena pada media sosial *twitter* terlalu banyak bahasa yang kurang baku.

DAFTAR PUSTAKA

Agarwal, A. et al. (2014). Sentiment Analysis Of Twitter Data. *Departement Of Computer Science Columbia University*, Avaible at:http://www.cs.colombia.edu/~julia/papers/Agarwaleta11.pdf.

Coletta, L. e. (2014). Combining Classification and Claustering for Tweet Sentimen Analysis. In 2014 Brazilian Conference on Interlligent systems.Avaliable.*at:http://ieeexplore.ieee.prg/lpdocs/epic03/wrapper:thm?anumber=6984832[Accessed May,11,2015]*, IEE,pp 210-215.

Ghulam Asrofi, B. (2017). *Analisis Sentimen Calon Gubernur DKI Jakarta 2017 DI Twitter*. Universitas Muhammadiyah Ponorogo.

Giummole, F. O. (2013). Trending Topics on Twitter Improve The Prediction of Google Hot Queries. In 2013 International Conference on Social Computing.IEEE,.*Avaible at:http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?anumber=6693309[AccessedJune7,2014*], PP39-44.

Http://setkab.go.id/inilah-undang-undang-nomor-7-tahun-2017-tentangpemilihan-umum-1/.(Di akses pada tanggal 15 April 2018).

Http://news.detik.com/berita/3928922/survei-pilgub-jatim-pltrackingvs-charta-berola-belakang. (Di akses pada tanggal 15 April 2018).

Http://www.viva.co.id/siapa/read/224-khofifah-indar-parawansa (Di akses pada tanggal 15 April 2018).

Http://www.viva.co.id/siapa/read/587-saifullah-yusuf (Di akses pada tanggal 15 April 2018).

Larasati, R. (2015). https://repository.widyatama.ac.id/xmlui/bitstream/handle/..../6830/bab02.pdf.

Liu, B. (2012)., Sentimen Analysis and Opiion Mining., Morgan & Claypool Publishers. Diakses dari https://www.cs.uic.edu/~liub/FBS/SentimenAnalysis-and OpinionMining.pdf.

Marfian , D. (2015). https://repository.widyatama.ac.id/xmlui/bitstream/handle/..../5668/bab2.pdf.

N.D Putranti and E. Winarko (2017) , Analisis Sentimen Twitter untuk Teks Berbahsa Indonesia dengan Maximum Entrophy dan Support Vector Machine, *IJCCS (Indonesian J. Compt, Cybern. Syt*.vol.8 no.1,pp.91 100,2014[Online].Availablle:htpps://jurnal.ugm.ac.id/ijccs/article/view/3499.

Noviantini. A., Sabariah, M.K &, E. (2015). *Analisis Sentimen*

pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine.

Pushpita Anna Octaviani, Y. W. (2014). Penerapan Metode Klasifikasi Support Vector Machine (SVM) pada data Akreditasi Sekolah Dasar (SD) Di Kabupaten Malang, Halaman 811-820.

Team, T. T. (2014). The Twitter Goverment and Election Handbook,. *San Francisco:TwitterInc.Available.at:https://g.twing.com/elections/files/2014/09/16/TwitterGovElectionsHamdbook.pdf.*

Top Media Sosial http://www.evadollzz.com/2014/09/top-10-socialnetworkingsterpopuler.html (2014).

Zainuddin, Nurulhuda dan Ari Selamat (2014). Sentimen Analysis Using Support Vector Machine. *Internantional Conference on Computer, Communication, and Control Technology*.12.

UNIVERSITAS MUHAMMADIYAH
JEMBER