

ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PARTAI POLITIK PENDUKUNG CALON GUBERNUR DI JAKARTA MENGGUNAKAN ALGORITMA C4.5 DECISION TREE LEARNING

Muhammad Rizqo Fajriansyah¹⁾, Ir Siswanto, M.M²⁾

^{1,2} Program studi, Fakultas Teknologi Informasi – Universitas Budi Luhur
Jl. Raya Ciledug. Petukangan Utara, Kebayoran Lama, Jakarta Selatan 12260
E-mail : rizqo.fajriansyah@gmail.com¹⁾, siswanto@budiluhur.ac.id²⁾

ABSTRAK

Internet telah menjadi sarana berbagi informasi dan mengekspresikan opini dengan respon yang cepat oleh pengguna di jejaring sosial. Layanan jejaring sosial memungkinkan setiap orang saling berinteraksi dan berbagi informasi tanpa batasan ruang dan waktu. Salah satu jejaring sosial yang populer dan banyak digunakan adalah Twitter. Pada tahun 2017 dilaksanakan pemilihan calon gubernur yang menjadi perhatian masyarakat, informasi digunakan masyarakat untuk menilai partai politik peserta salah satunya opini negatif dan positif yang berasal dari twitter. Dalam menentukan polaritas positif atau negatif suatu opini dapat dilakukan secara manual, tetapi mempertimbangkan bertambahnya opini menjadi semakin banyak, tentunya banyak waktu yang akan semakin banyak terpakai. Maka diajukan sebuah metode *machine learning* untuk mengklasifikasikan konten opini dari sumber data yang sangat banyak. Metode *machine learning* untuk melakukan analisis terhadap konten *tweet* yaitu menggunakan pohon keputusan atau *decision tree learning* yang dibangun dengan algoritma C 4.5. Pada penelitian ini akan dilakukan analisis *tweet* yang bersifat opini terhadap partai politik peserta pemilihan calon gubernur 2017. Analisis sentimen terhadap partai politik dilakukan untuk menganalisa opini negatif dan positif dari pengguna twitter terhadap partai peserta pemilihan calon gubernur 2017. Hasil dari pengujian alpha menunjukkan bahwa sistem analisis sentimen menggunakan algoritma C4.5 bebas dari kesalahan sintak dan berjalan sesuai dengan fungsinya. Hasil pengujian aplikasi dapat disimpulkan bahwa aplikasi analisis sentimen ini baik untuk referensi tambahan masyarakat terhadap partai politik. Hasil dari pengujian akurasi klasifikasi menggunakan pohon keputusan yang dibangun algoritma C4.5 berdasarkan data latih dan 150 data uji yang digunakan menghasilkan rata rata *error* 36%.

Kata Kunci : Sentimen, Partai Politik, *Text Processing*, Algoritma C.4.5.

1. PENDAHULUAN

Internet telah menjadi sarana berbagi informasi dan mengekspresikan opini dengan respon yang cepat oleh pengguna di jejaring social, terhadap topik yang bermacam-macam dalam bentuk 140 Karakter pesan yang dinamakan *tweet*, twitter juga dikenal sebagai sarana penyebaran informasi dan mengekspresikan opini . Contohnya selama peristiwa besar terjadi seperti bencana tsunami, gunung meletus, badai ataupun peristiwa seperti pemilihan umum, ataupun pendapat terhadap tokoh. Hal-hal tersebut selalu ditanggapi oleh aktivitas di internet . Pengguna twitter menggunakan sarana ini untuk berpartisipasi mengekspresikan opininya. Pengguna twitter menyebarkan *tweet* dengan konten yang berbeda-beda sesuai dengan pola kebiasaan dan gaya menulis pengguna. Konten seperti ini dapat mengandung informasi yang bernilai. Konten yang disebarkan pengguna twitter dapat dikategorikan, secara umum kategorinya yaitu informasi/berita, opini/keluhan, promosi, pernyataan, *random tweet*, pertanyaan, membuktikan eksistensi, obrolan dan *tweet* mengenai pengguna tersebut . Pada penelitian ini akan dilakukan analisis *tweet* yang bersifat opini

terhadap partai politik peserta Pendukung calon gubernur Menggunakan Algoritma C4.5 *Decision Tree Learning*. Adapun tujuan yang ingin dicapai dari Analisis Sentimen Pengguna Twitter Terhadap Partai Politik Pendukung calon gubernur Menggunakan Algoritma C4.5 diantaranya adalah dapat mengetahui tingkat akurasi pohon keputusan yang dibentuk algoritma C4.5 dalam mengklasifikasikan *tweet* sentiment dan dapat menjadi referensi tambahan masyarakat terhadap partai politik.

Twitter adalah jejaring sosial dan *microblogging service* yang memungkinkan penggunanya untuk mengirim 140 karakter pesan teks yang disebut "*tweet*". Twitter memiliki popularitas yang mendunia dengan 29,5 juta pengguna yang terdaftar pada tahun 2012 yang mengirim sebanyak 340 juta *tweet* per hari

Salah satu contoh implementasi *text mining* termasuk mengklasifikasikan dokumen ke dalam topik yang spesifik atau *supervised learning*, mengelompokan dokumen yang mempunyai kemiripan atau *unsupervised learning*, dan diimplementasikan ke dalam sistem pencarian dokumen atau *information retrieval*. *Text data*

mining fokus terhadap metode *data mining* yang diaplikasikan pada tekstual data. Beberapa bagian dari *text mining* adalah *term*, *corpus*, *lexicon*, dan dokumen. *Term* adalah kata atau kombinasi kata atau prasa. Dokumen adalah sekumpulan kata dan tanda baca, dokumen dapat terdiri dari beberapa segmen teks dan memiliki panjang yang tidak ditentukan. Contoh dari dokumen adalah kalimat, paragraf, bagian sebuah buku atau bab, buku, halaman web, email dan lain-lain. *Corpus* adalah koleksi dokumen. *Lexicon* adalah sekumpulan kata yang bersifat unik yang terdapat pada *corpus*.

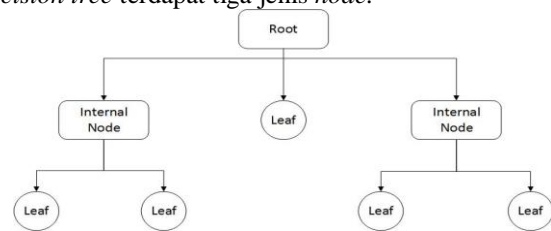
Text Pre-Processing adalah serangkaian proses yang dilakukan untuk mempersiapkan data sebelum proses lebih lanjut. Tahapan ini memiliki tujuan untuk mengubah data agar dapat dilanjutkan ke proses klasifikasi. Terdapat beberapa proses yang harus dilakukan dalam tahap *text pre-processing* atau persiapan dokumen teks, yaitu *tokenization*, *casefolding*, *stemming*. *Tokenization* merupakan proses memecah suatu kalimat dalam seluruh isi dokumen menjadi sekumpulan kata atau token yang dipisahkan oleh spasi. Hasil dari proses *tokenization* selanjutnya digunakan pada proses selanjutnya seperti *casefolding* dan *stemming*. *Case Folding*, merupakan proses mengubah semua huruf yang ada pada seluruh dokumen menjadi huruf kecil. Huruf "a" sampai dengan "z" yang akan diproses, karakter selain huruf akan dianggap *delimiter*. *Stemming* adalah salah satu cara untuk meningkatkan performa dalam *text processing* dengan cara mentransformasi kata-kata dalam sebuah dokumen teks ke kata dasarnya. Algoritma *stemming* untuk masing-masing bahasa berbeda-beda, sebagai contoh bahasa Inggris memiliki morfologi yang berbeda dengan bahasa Indonesia sehingga algoritma *stemming* untuk kedua bahasa tersebut juga berbeda.

Salah satu aplikasi dari data mining adalah klasifikasi. Menggunakan sampel data sebelumnya yang telah memiliki kategori untuk dijadikan data training, tujuannya adalah untuk memberikan kategori pada data baru yang belum dikategorikan. Klasifikasi pada teks melibatkan banyak teknik termasuk *indexing* pada *information retrieval* dan teknik dari *machine learning* untuk menentukan kategori secara otomatis dari data training yang sudah diberi label. Pada proses klasifikasi sebuah dokumen, tidak digunakan informasi lain selain konten dari dokumen tersebut.

Banyak teknik *machine learning* yang digunakan untuk klasifikasi diantaranya adalah *Naïve Bayesian*, *Decision Tree*, *k-Nearest Neighbor* (k-NN), *Neural Network* (NN), dan *Support Vector Machines* (SVM). Teknik teknik tersebut berbeda dalam mekanisme pembelajaran dan representasi model yang dipelajari.

Mekanisme dari pohon keputusan seperti struktur pohon, dimana tiap *internal node* menunjukkan sebuah test pada sebuah atribut, setiap cabang menunjukkan hasil dari test, dan *leaf node*

menunjukkan kelas atau kategori. Pada gambar 1 *decision tree* terdapat tiga jenis *node*.



Gambar 1. Pohon Keputusan

Algoritma C4.5 dibangun oleh Ross Quinlan merupakan ekstensi atau pengembangan dari algoritma ID3. Algoritma ini merupakan kelompok algoritma pohon keputusan. Mempunyai masukan berupa data training. Data training berupa data contoh yang telah mempunyai label kelas positif atau negatif, akan digunakan untuk membangun sebuah *tree*. Sedangkan atribut-atribut data digunakan sebagai parameter dalam proses klasifikasi data.

K-fold cross-validation adalah sebuah pengujian dimana *dataset X* dibagi secara acak kedalam *K* bagian dengan ukuran yang sama, $X_i, i=1, \dots, K$. Untuk memproses tiap pasangannya, diambil terlebih dahulu salah satu bagian dari *K* bagian sebagai data tes, dan mengkombinasikan sisanya sebanyak *K-1* bagian sebagai data latih. Proses pengujian dilakukan sebanyak *K* kali.

Dari pengujian yang dilakukan, dievaluasi dengan menggunakan *confusion matrix* untuk mengetahui tingkat kebenaran dari penelitian seperti Tabel 1.

Tabel 1. *Confusion Matrix*

Predicted Class	Tested Class		
	Positive	Negative	Total
Positive	True Positive	False Positive	P
Negative	False Negative	True Negative	
Total	p'	n'	N

2. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini, langkah-langkah sebagai berikut:

- Analisa Kebutuhan dilakukan dengan penelitian langsung ke PT Comindo Unggul Lestari yang diteliti untuk mendapatkan data dan informasi yang harus diamankan serta masalah yang sering terjadi selama proses transfer file yang berisikan data penting.
- Mempelajari cara kerja algoritma C45 dan *Decision Tree*.
- Mendesain serta memodelkan algoritma dan user interface aplikasi Analisis Sentimen Pengguna Twitter.
- Membuat program dengan bahasa pemrograman PHP dan mengelola file log proses text mining dengan MySQL

- e. Ujicoba Program dengan mencoba panjang pesan chat yang kecil dan yang besar ukuran pesannya.

3. HASIL DAN PEMBAHASAN

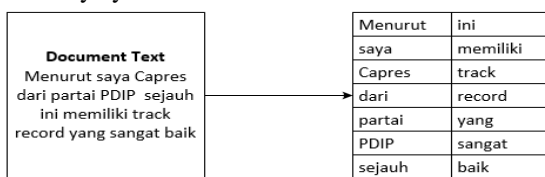
Dalam proses pembuatan suatu sistem dilakukan analisis terhadap sistem yang akan dibangun. Analisis yang dilakukan untuk membangun aplikasi analisis sentimen pengguna twitter terhadap partai politik peserta PEMILU 2014 menggunakan pohon keputusan yang dibangun oleh algoritma C4.5.

3.1 Proses Text Pre-processing

Proses *Text Pre-processing* dilakukan untuk mempersiapkan data sebelum ke proses selanjutnya, yaitu pemberian kelas kata untuk setiap kata dalam suatu *tweet* dan klasifikasi sentimen positif dan negatif. Terdapat beberapa sub proses dari *text pre-processing*, yaitu *tokenization*, *case folding*, dan *stemming*.

3.2 Tokenization

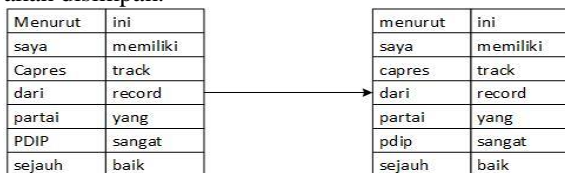
Tokenization dilakukan untuk memecah suatu kalimat dalam seluruh isi dokumen menjadi sekumpulan kata. Setiap kata yang telah dipecah dari kalimat/*tweet* akan disimpan ke basis data dengan atributnya yaitu id *tweet*.



Gambar 2. *Tokenization*

3.3 Casefolding

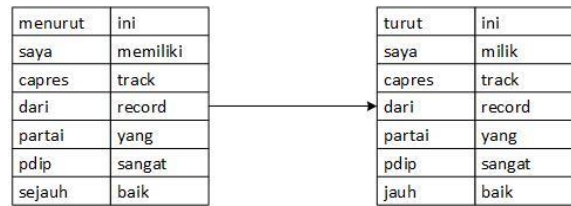
Case folding adalah proses mengubah semua huruf yang ada pada seluruh dokumen menjadi huruf kecil. Huruf “a” sampai dengan “z” yang akan diproses, karakter selain huruf akan dianggap *delimiter*. Proses ini dilakukan untuk menghindari duplikasi data sehingga kata yang memiliki perbedaan huruf kecil dan besar tetapi sama maka kata tersebut tidak akan disimpan.



Gambar 3. *Casefolding*

3.4 Stemming

Algoritma *stemming* yang digunakan pada penelitian ini adalah algoritma nazief dan adriani yang dibuat oleh Bobby Nazief dan Mirna Adriani



Gambar 4. *Stemming*

3.5 Pembentukan Pohon Keputusan

Data hasil *text preprocessing* selanjutnya diklasifikasikan menggunakan pohon keputusan yang dibangun algoritma C4.5. Sebelum pohon keputusan digunakan untuk mengklasifikasikan ke data sebenarnya, model pohon keputusan belajar dari data *training* yang telah mempunyai kelas, yaitu negatif dan positif.

Pada seluruh dokumen *tweet* yang ada, setiap kata yang muncul dijadikan atribut untuk perhitungan entropi dan *gain*. nilai dari atribut tersebut adalah biner (0 dan 1)., dimana nilai 0 mengartikan tidak munculnya kata tersebut pada *tweet*, dan nilai 1 mengartikan bahwa kata tersebut muncul pada *tweet*.

3.6 Perhitungan Information Gain

Berikut ini adalah contoh proses pembentukan pohon keputusan menggunakan algoritma C4.5 menggunakan empat kalimat opini yang diberi kelas negatif dan positif mengenai PEMILU.

Tabel 2. Data Latih

No.	Kalimat Opini	Kelas
1	Pemilu gagal	Negatif
2	Pemilu gagal dicurangi	Positif
3	Pemilu sukses	Positif
4	Pemilu dicurangi	Negatif

Proses perhitungan gain masing-masing atribut terhadap training data dilakukan untuk menentukan atribut pengujian pada setiap simpul. Didapatkan empat kata yang akan dijadikan atribut pengujian setelah dilakukan proses *text preprocessing* yaitu “pemilu”, “curang”, “sukses”, dan “gagal”. Selanjutnya setiap sampel dimasukan kedalam sebuah tabel dengan keempat atribut pengujian, atribut pengujian bernilai 1 jika sampel mengandung kata pada atribut pengujian dan bernilai 0 jika tidak mengandung kata pada atribut pengujian. Berikut ini tabel yang dihasilkan dari training data dan proses perhitungan information gain terhadap semua atribut pengujian.

Tabel 3. Nilai Data Latih

No	curang	pemilu	gagal	sukses	kelas
1	0	1	1	0	Negatif
2	1	1	1	0	Negatif
3	0	1	0	1	Positif
4	1	1	0	0	Negatif

Perhitungan entropi untuk seluruh training data

$$Entropy(Total) = -\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4} = 1$$

Perhitungan information gain pada setiap atribut pengujian untuk menentukan simpul akar :

a curang

Perhitungan entropy atribut “curang” dengan nilai 0 dan 1:

$$Entropy(0) = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1$$

$$Entropy(1) = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1$$

$$Gain(Total, curang) = 1 - ((2/4)$$

$$Entropy(0) + (2/4) Entropy(1)) = 0$$

b pemilu

Perhitungan entropy atribut “pemilu” dengan nilai 0 dan 1:

$$Entropy(0) = -0 - 0 = 0$$

$$Entropy(1) = -\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4} = 1$$

$$Gain(Total, pemilu) = 1 - ((0/4) Entropy(0)$$

$$+ (4/4) Entropy(1)) = 0$$

c gagal

Perhitungan entropy atribut “gagal” dengan nilai 0 dan 1:

$$Entropy(0) = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1$$

$$Entropy(1) = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1$$

$$Gain(Total, gagal) = 1 - ((2/4) Entropy(0)$$

$$+ (2/4) Entropy(1)) = 0$$

d sukses

Perhitungan entropy atribut “sukses” dengan nilai 0 dan 1:

$$Entropy(0) = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} = 0.918$$

$$Entropy(1) = -\frac{1}{1} \log_2 \frac{1}{1} - 0 = 0$$

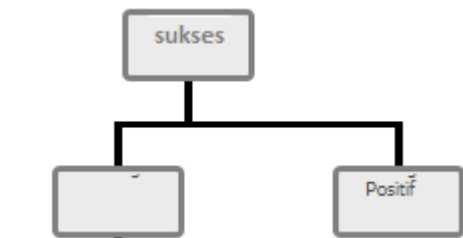
$$Gain(Total, sukses) = 1 - ((3/4) Entropy(0)$$

$$+ (1/4) Entropy(1)) = 0.312$$

Tabel 4. Perhitungan Information Gain Untuk Simpul Akar

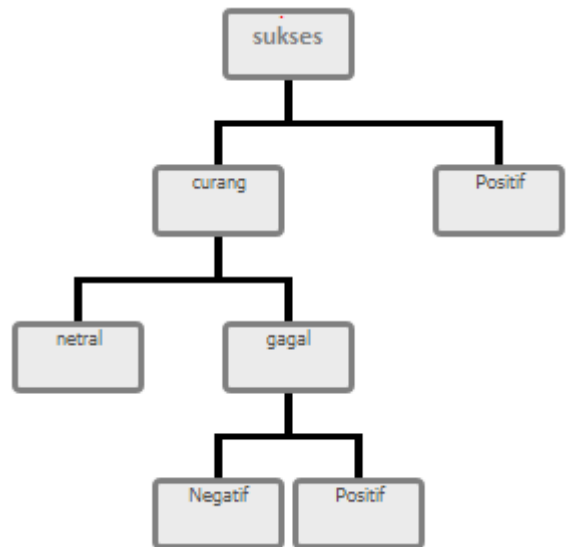
Perhitungan Node Akar							
Total Entropy : 1							
Attr ibut	Ne g(0)	Po s(0)	Entro py(0)	Ne g(1)	Po s(1)	Entro py(1)	Ga in
pem ilu	0	0	0	2	2	1	0
gag al	1	1	1	1	1	1	0
cura ng	1	1	1	1	1	1	0
suk ses	2	1	0.918	0	1	0	0. 31 2
Max Gain : 0.312							
Node : sukses							

Hasil perhitungan nilai gain untuk menentukan simpul akar akan membentuk *rule tree* seperti pada gambar.



Gambar 5. Hasil Perhitungan Nilai Gain Simpul Akar

Setelah pembentukan pohon keputusan berakhir pohon keputusan yang dibentuk berdasarkan data latih seperti pada gambar.



Gambar 6. Pohon Keputusan Akhir

Pohon keputusan yang dibangun oleh algoritma C4.5 ini bergantung pada training data yang disiapkan. *Tweet* yang akan diklasifikasikan akan diuji dimulai dari simpul teratas atau simpul akar. Jika sebuah *tweet* mengandung kata yang menjadi atribut pengujian pada simpul akar yaitu ‘sukses’ maka kata atau simpul selanjutnya yang akan diuji adalah simpul ruas kanan, jika *tweet* tidak mengandung kata ‘sukses’ maka simpul yang diuji selanjutnya adalah simpul ruas kiri, proses ini akan terus dilakukan sampai ditemukan simpul daun yang berisi kelas sentimen negatif, positif dan netral. Jika sebuah *tweet* mengandung kata yang tidak diuji pada jalur yang sedang ditelusuri, maka kata tersebut tidak akan diuji.

3.7 Klasifikasi Menggunakan Pohon Keputusan

Model pohon keputusan yang telah dibentuk menggunakan algoritma C4.5 selanjutnya digunakan untuk mengklasifikasikan sebuah kalimat opini. Kata-kata yang akan diuji adalah kata-kata yang menjadi atribut pengujian pada pohon keputusan, jika pada kalimat opini terdapat kata yang tidak menjadi atribut pengujian pada pohon keputusan maka kata tersebut tidak akan diuji. Jika pada kalimat terdapat kata yang menjadi atribut pengujian simpul yang sedang diuji maka simpul berikutnya yang diuji

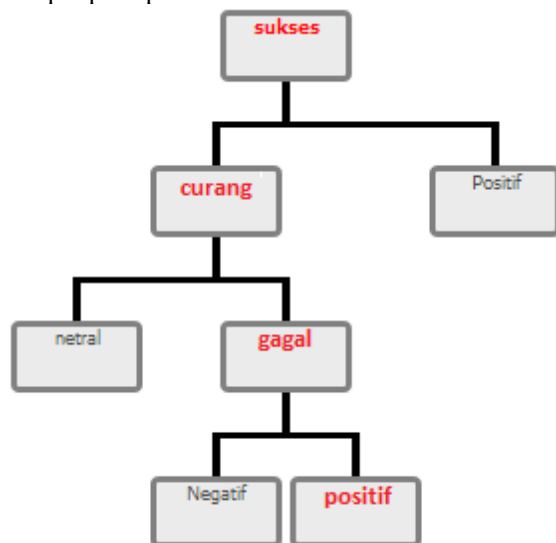
adalah simpul ruas kanan dan seluruh simpul ruas kiri tidak akan diuji begitu juga sebaliknya. Berikut adalah langkah-langkah klasifikasi menggunakan pohon keputusan :

- Simpul akar akan menjadi simpul pertama yang akan diuji,
- jika pada kalimat terdapat kata yang menjadi atribut pengujian pada simpul maka selanjutnya akan diuji simpul pada ruas kanan yang merupakan cabang dari simpul yang diuji sebelumnya. Jika pada kalimat tidak terdapat kata yang menjadi atribut pengujian pada simpul tersebut maka selanjutnya akan diuji simpul pada ruas kiri yang merupakan cabang dari simpul yang diuji sebelumnya.
- Jika pada simpul yang diuji bukan merupakan *leaf node* maka akan dilakukan langkah sebelumnya yaitu langkah 2 pada simpul yang sedang diuji.
- Proses akan berhenti jika simpul yang diuji merupakan *leaf node*.

Selanjutnya adalah contoh klasifikasi sebuah kalimat menggunakan pohon keputusan yang telah dibentuk.

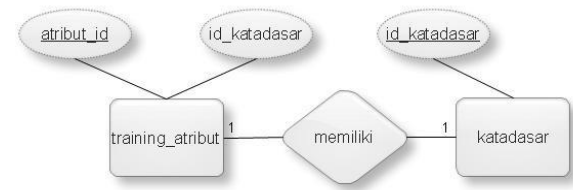
Kalimat Opini : *“Pemilu gagal dicurangi”*

Simpul pertama yang akan dilakukan pengecekan adalah simpul kata “sukses” yang tidak terdapat pada kalimat maka simpul selanjutnya yang dilakukan pengecekan adalah simpul pada ruas kiri yaitu curang. Kata “curang” terdapat pada kalimat sehingga pengecekan selanjutnya dilakukan pada simpul ruas kanan yaitu “gagal”. Kata “gagal” terdapat pada kalimat sehingga pengecekan selanjutnya dilakukan pada simpul ruas kanan yaitu “Positif”. Karena pengecekan simpul telah mencapai simpul daun, maka pengecekan dihentikan dan menghasilkan “Positif” sebagai kelas dari kalimat yang diuji. Berikut ini adalah gambar pengecekan simpul pada proses klasifikasi.



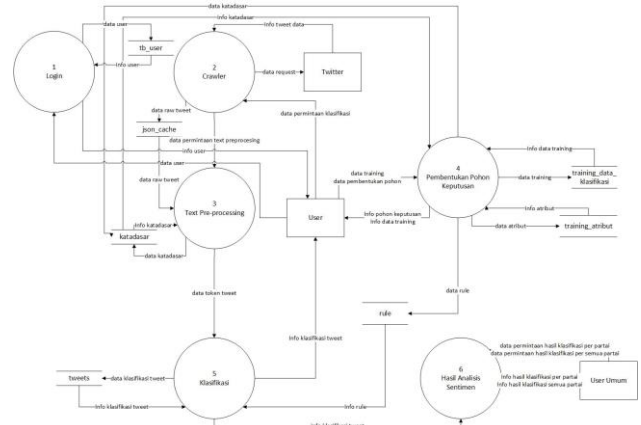
Gambar 7. Proses Pengecekan Kalimat Opini

3.8 ERD



Gambar 15. ERD yang digunakan pada Sistem Analisis Sentimen Pengguna Twitter Terhadap Partai Politik

3.9 DFD



Gambar 16. DFD Level 1 Analisis Sentimen Pengguna Twitter Terhadap Partai Politik

Implementasi dari sistem dibangun berbasis website.

a. Tampilan Halaman Login

Menu tampilan halaman login ditampilkan pada saat awal user akan membuka tampilan kelola, sebelum user dapat menggunakan fitur secara penuh diperlukan akses bagi user tersebut.



Gambar 17. Implementasi Halaman Login

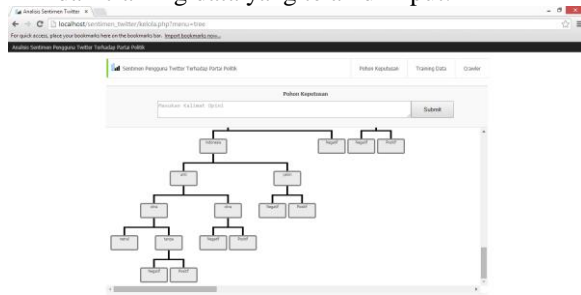
b. Tampilan Halaman User

Menu tampilan halaman user menampilkan persentase dari hasil tweet yang sudah disaring dan dibuat berbentuk graphic chart dengan tampilan positif dan negative.



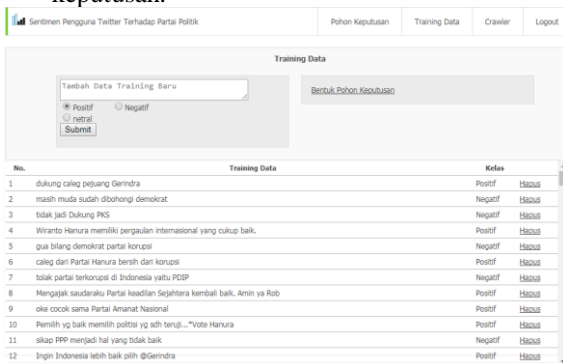
Gambar 18. Implementasi Halaman User

- c. Tampilan Halaman Pohon Keputusan
Menu tampilan pohon keputusan menampilkan pohon keputusan yang dibentuk berdasarkan dari training data yang telah di input.



Gambar 19. Implementasi Halaman Pohon Keputusan

- d. Tampilan Halaman Training Data
Menu tampilan halaman training data berisi statement dengan kalimat yang mengandung nilai entropi untuk diubah menjadi pohon keputusan.



Gambar 20. Implementasi Halaman Training Data

- e. Tampilan Halaman Tweet Partai
Menu tampilan halaman tweet partai berisi tweet yang telah tersaring dan memenuhi kriteria untuk ditampilkan di dalam tampilan utama.



Gambar 18. Implementasi Halaman Tweet Mengenai Partai

4.1 Pengujian Pohon Keputusan

Pengujian ditujukan untuk mengukur tingkat kesesuaian antara hasil klasifikasi terhadap data sentimen secara manual dengan hasil klasifikasi menggunakan pohon keputusan yang dibentuk menggunakan algoritma C4.5. Adapun pengujian yang dilakukan menggunakan k-fold cross validation dengan nilai k = 3. Nilai 3 menunjukkan jumlah fold data, dimana masing – masing subset memiliki data yang berbeda. Iterasi pengujian akan dilakukan sebanyak tiga kali.

Jumlah data yang digunakan sebanyak 150 data, data ini merupakan data yang telah dikelompokkan secara manual dengan jumlah sebesar 150 data untuk masing-masing kelasnya. Tiap fold terdiri dari 50 data.

Tabel 6. Kombinasi Fold

Iterasi	Data Uji	Data Latih
1	F1	F2, F3
2	F2	F1, F3
3	F3	F1, F2

1. Iterasi Satu

Digunakan 100 data sebagai data latih dan 50 data sebagai data uji. Adapun data latih yang digunakan adalah fold 2, fold 3, sedangkan data yang akan diuji adalah fold 1. Hasil pengujian fold pertama dengan menggunakan fold cross validation seperti dapat dilihat pada tabel berikut.

Tabel 7. Cross Validation Fold 1

Hasil Klasifikasi Manual	Hasil Pengujian Class		
	Positif	Negatif	Total
Positif	18	13	31
Negatif	5	13	19
Total	27	23	50

2. Iterasi Dua

Digunakan 100 data sebagai data latih dan 50 data sebagai data uji. Adapun data latih yang digunakan adalah fold 1, fold 3, sedangkan data yang akan diuji adalah fold 2. Hasil pengujian fold kedua dengan menggunakan fold cross validation seperti dapat dilihat pada tabel berikut.

Tabel 8. Cross Validation Fold 2

Hasil Klasifikasi Manual	Hasil Pengujian Class		
	Positif	Negatif	Total
Positif	16	12	28
Negatif	8	14	22
Total	24	26	50

3. Iterasi Tiga

Digunakan 100 data sebagai data latih dan 50 data sebagai data uji. Adapun data latih yang digunakan adalah *fold 1*, *fold 2*, sedangkan data yang akan diuji adalah *fold 3*. Hasil pengujian *fold* ketiga dengan menggunakan *fold cross validation* seperti dapat dilihat pada tabel berikut.

Tabel 9. *Cross Validation* Fold 3

Hasil Klasifikasi Manual	Hasil Pengujian Class		
	Positif	Negatif	Total
Positif	20	12	32
Negatif	4	14	18
Total	24	26	50

Setelah dilakukan iterasi sebanyak tiga kali, didapat hasil pengujian seperti dapat dilihat pada tabel berikut.9

Tabel 10. Nilai *Error* Pengujian

Data Uji	Hasil Pengujian
	Nilai Error
Iterasi 1	36%
Iterasi 2	40%
Iterasi 3	32%
Rata-rata Error	36%

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan mengenai analisis sentimen pengguna twitter terhadap partai politik menggunakan algoritma C4.5, maka dapat disimpulkan bahwa:

Berdasarkan hasil pengujian menggunakan data latih dan data uji digunakan, maka diperoleh kesimpulan bahwa sistem diketahui tingkat rata-rata *error* yang diberikan pohon keputusan yang dibangun algoritma C4.5 dalam mengklasifikasikan kalimat sentimen ke dalam kelas positif dan negatif sebesar 36%.

Agar aplikasi analisis sentimen pengguna twitter terhadap partai politik menggunakan algoritma C4.5 ini semakin baik untuk kedepannya, maka ada beberapa saran yang perlu dilakukan dalam penelitian berikutnya, diantaranya:

- a Pohon keputusan yang dibangun harus berdasarkan training data yang selengkap mungkin.
- b Kata kata pada kalimat opini yang menjadi training data disarankan untuk meminimalkan kata kata yang tidak perlu agar meminimalkan munculnya kata-kata penguji yang tidak perlu pada pohon keputusan.

DAFTAR PUSTAKA

[1] Gupta, A., P. Kumaraguru. "Credibility Ranking of Tweets during High Impact Events". In *Proceedings of the 1st Workshop on Privacy and Security in Online Social Media*, vol 12, pp. , Apr.2012.

[2] Gupta, A., Lamba, H., and Kumaraguru, P. "\$1.00 per RT#BostonMarathon #PrayForBoston: Analyzing Fake Content on Twitter." *IEEE APWG eCrime Research Summit (eCRS)*, 2013.

[3] H. Christopher."Analysis and Classification of Twitter Messages." M. Thesis, Graz University of Technology, Austria, 2010.

[4] W. Xindong., "Top 10 Algorithms in Data Mining." *Knowledge and Information Systems*, vol 14, pp.1-37, Dec. 2007.

[5] Puranti, N.D., Winarko, E. "Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine". *IJCC*, vol 9, pp 91-100, 2014.

[6] L. S. Jeffrey., "Text Data Mining: Theory and Method". *Statistics Surveys*, vol. 2, pp. 94-112, Jul. 2008.

[7] A. Ledy. "Perbandingan Algoritma Stemming Porter dengan Algoritma Nazief & Adriani untuk Stemming Dokumen Teks Bahasa Indonesia". *Konferensi Nasional Sistem dan Informatika*, 2009.

[8] A. Aijun., "Classification Methods". *Encyclopedia of Data Warehousing and Mining, Second Edition*, vol 4, pp. 196-201, 2009.