

**ANALISIS SENTIMEN PADA MEDIA ONLINE DALAM PEMILIHAN
PRESIDEN 2019 DENGAN METODE DECISION TREE C4.5**

Tugas Akhir

diajukan untuk memenuhi salah satu syarat

memperoleh gelar sarjana

dari Program Studi Ilmu Komputasi

Fakultas Informatika

Universitas Telkom

1302154127

Musyafa'noer Sandi Pratama



Program Studi Sarjana Ilmu Komputasi

Fakultas Informatika

Universitas Telkom

Bandung

2019

LEMBAR PENGESAHAN

**Analisis Sentimen pada Media *Online* dalam Pemilihan Presiden 2019 dengan Metode
*Decision Tree C4.5***

***Sentiment Analysis on Online Media in 2019 Presidential Election with C4.5
Decision Tree Method***

**NIM :1302154127
Musyafa'noer Sandi Pratama**

Tugas akhir ini telah diterima dan disahkan untuk memenuhi sebagian syarat memperoleh gelar
pada Program Studi Sarjana Ilmu Komputasi
Fakultas Informatika
Universitas Telkom

Bandung, 23 Maret 2019

Menyetujui

Pembimbing,



Dr. Kemas Muslim L., S.T., M.ISD.

NIP:13820075

Ketua Program Studi
Sarjana Ilmu Komputasi,



Dr. Deni Saepudin S.Si.,M.Si

NIP: 99750013

LEMBAR PERNYATAAN

Dengan ini saya, Musyafa'noer Sandi Pratama, menyatakan sesungguhnya bahwa Tugas Akhir saya dengan judul Analisis Sentimen pada Media *Online* dalam Pemilihan Presiden 2019 dengan Metode *Decision Tree C4.5* beserta dengan seluruh isinya adalah merupakan hasil karya sendiri, dan saya tidak melakukan penjiplakan yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan. Saya siap menanggung resiko/sanksi yang diberikan jika di kemudian hari ditemukan pelanggaran terhadap etika keilmuan dalam buku TA atau jika ada klaim dari pihak lain terhadap keaslian karya,

Bandung, 23 Maret 2019

Yang Menyatakan



Musyafa'noer Sandi Pratama

Sentimen Analisis pada Media *Online* dalam Pemilihan Presiden 2019 dengan Metode *Decision Tree C4.5*

Musyafa'noer Sandi Pratama¹, Kemas Muslim L.²

^{1,2} Fakultas Informatika, Universitas Telkom, Bandung

¹emusyafa@students.telkomuniversity.ac.id, ²kemasmuslim@telkomuniversity.ac.id

Abstrak

Analisis sentimen adalah riset komputasional dari opini sentiment dan emosi yang diekspresikan secara tekstual. Pada pemilihan presiden 2019 ini akan banyak opini di situs berita tentang pasangan capres dan cawapres. Pada opini tersebut ada yang condong hanya pada satu pasangan capres dan cawapres atau bisa disebut positif dan negatifnya opini. Ada banyak berita pada media *online* mengenai pemilihan presiden 2019 ini, karena itu dibutuhkan klasifikasi pada berita di media *online* agar menjadi lebih mudah mengetahui kecondongan media *online* tersebut pada salah satu capres dan cawapres. Pada penelitian ini akan diimplementasikan metode klasifikasi *decision tree C4.5* pada teks atau dokumen mengenai pemilihan presiden 2019 untuk mengetahui sentimen positif atau negatif dan akurasi. Dengan cara mengambil data uji menggunakan *web scraping* pada situs berita *online*.

Kata Kunci : Sentimen Analisis, *Decision Tree C4.5*, *Web Scraping*, *Media Online*.

Abstrak

Sentiment analysis is a computational research of sentiment and emotion opinions expressed textually. In the 2019 presidential election there will be many opinions on news sites about presidential and vice presidential couples. In that opinion there are those who are inclined to only one presidential and vice presidential couple or can be called positive and negative opinions. There is a lot of news on the online media regarding this 2019 presidential election, because it requires classification on the news in online media so that it becomes easier to know the bias of the online media in one of the presidential and vice presidential candidates. In this study C4.5 decision tree classification method will be implemented in the text or document regarding 2019 presidential elections to find out positive or negative sentiments and accuracy. By taking test data using web scraping on online news sites.

Keywords : *Sentiment Analysis, Decision Tree C4.5, Web Scraping, Media Online.*

1. Pendahuluan

1.1. Latar Belakang

Berita adalah sebuah informasi tentang apa yang sedang atau telah terjadi di suatu tempat baik berupa cetak, siaran, internet atau dari mulut ke mulut. Selain sarana penyampai informasi, berita juga merupakan penggiring opini publik yang cukup efektif. Semakin dekatnya pemilihan presiden dan wakil presiden 2019 membuat masyarakat mulai terpecah kedalam 2 kubu yaitu kubu Jokowi dan Prabowo. Penyedia layanan berita yang harusnya bersifat netral, beberapa malah menjadi senjata dalam menggiring opini publik ke suatu kubu.

Analisis ini menggunakan media *online* sebagai media untuk mengumpulkan data berupa berita yang sifatnya netral atau condong ke salah satu pasangan capres dan cawapres dengan metode sentimen analisis. Salah satu contoh dapat dilihat pada media detik.com dengan berita yang berjudul "Ke Pasar Ikan di Aceh, Sandiaga: Ekonomi Sepi, Pedagang Butuh Modal" dan berita dengan judul "Jokowi Jadi Raja, Erick Thohir: Saya Mundur" dari media tribunnews.com. Kedua judul tersebut akan dianalisis untuk mengetahui polaritas dari media tersebut.

Untuk mengetahui negatif atau positifnya suatu berita, maka dibutuhkan metode analisis sentimen terhadap berita-berita tersebut. Pada penelitian sebelumnya yang berkaitan dengan sentimen analisis pemilu [7], metode yang digunakan adalah klasifikasi *Support Vector Machine* (SVM) dan *Naïve Bayes Classifier* (NBC). Hasil Nilai akurasi tertinggi didapat saat menggunakan metode klasifikasi *Naïve Bayes Classifier* (NBC), dengan nilai rata-rata akurasi mencapai 95%, nilai presisi 95%, nilai recall 95% nilai TP rate 96,8% dan nilai TN rate 84,6% [7].

Pada Tugas Akhir ini penulis menggunakan metode klasifikasi *Decision Tree* karena ingin membuktikan bahwa metode ini lebih efektif, misal suatu data yang kompleks menjadi lebih simpel dan penelitian yang lain pun mengatakan *Decision Tree* dapat bekerja lebih baik dari metode klasifikasi lainnya [5,7].

1.2. Tujuan

Tujuan pembuatan Tugas Akhir ini adalah.

1. Menghasilkan sistem yang dapat menentukan sebuah teks berita memiliki sentimen positif atau negatif kepada salah satu kubu, Jokowi-Ma'ruf atau Prabowo-Sandi menggunakan metode *decision tree* C4.5.
2. Menguji dan menganalisis performansi klasifikasi *decision tree* C4.5 untuk menentukan kumpulan berita mempunyai sentimen positif atau negatif ke kubu Jokowi-Ma'ruf atau kubu Prabowo-Sandi.

1.3. Batasan Masalah

1. Teks berita yang akan diolah hanya teks berbahasa Indonesia.
2. Teks berita yang di analisis adalah teks berita yang diterbitkan dimedia *online* pilpres.tempo.co dan hanya teks berita yang mengandung nama calon persiden dan wakil presiden baik dari kubu Jokowi atau Prabowo.
3. Tugas Akhir ini akan memfokuskan pada hasil klasifikasi yang dihasilkan dengan menggunakan metode *Decision Tree* C4.5.
4. Klasifikasi teks berita hanya kedalam positif dan negatif.
5. Berita yang diambil mulai dari 1 Desember 2018 – 28 Februari 2019.

2. Studi Terkait

2.1. Sentiment Analysis

Sentiment Analysis merupakan proses mengolah data untuk mendapatkan teks yang terdapat dalam di dalam suatu kalimat. *Sentiment Analysis* dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif [1].

2.2. Preprocessing

Preprocessing adalah suatu proses pengubahan bentuk data teks yang tidak terstruktur menjadi bentuk yang terstruktur sesuai dengan kebutuhannya agar data yang diambil mudah untuk dikelompokkan, biasanya data diubah kedalam bentuk *term*. *Term* adalah kata yang berdiri sendiri atau frasa (gabungan kata) yang diambil dalam dokumen. Pada umumnya *Preprocessing* memiliki beberapa tahapan proses yaitu,

1. Case Folding

Case folding adalah mengubah semua huruf yang berada pada teks menjadi huruf standar (biasanya jadi huruf kecil). Hanya karakter atau huruf 'a' sampai dengan 'z' yang akan diterima, selain itu akan dihilangkan dari dokumen. Sebagai contoh bila user ingin mencari informasi "CIWIDEY" dan mengetiknya "Ciwidey", atau "CiwidEY" tetap diberikan hasil yang sama yakni "ciwidey".

2. Tokenizing

Tokenizing adalah adalah proses pemecahan string input berdasarkan tiap kata penyusunnya. Pada prinsipnya dari sebuah kalimat dibuat menjadi perkata (satuan kata) dan menghilangkan karakter-karakter yang dianggap sebagai pemisah kata (delimiter) dan juga angka, misalnya tanda tanya (?), koma (,), titik (.), dsb. Pada proses ini juga dilakukan *cleaning* yaitu membersihkan dokumen dari komponen-komponen yang tidak terkait dengan informasi yang ada pada dokumen, seperti html, link, url, dsb.

3. Filtering

Filtering adalah proses pemilihan kata-kata penting dari hasil *tokenizing*. Untuk menghilangkan kata-kata tidak penting bisa menggunakan algoritma *stoplist* atau bisa juga menggunakan algoritma *worldlist* untuk menyimpan kata-kata yang penting. Dengan menggunakan algoritma *stoplist* kata-kata yang tidak deskriptif seperti "ke", "di", "akan", dan semacamnya dapat dibuang dengan pendekatan *bag-of-words*.

4. Stemming

Stemming adalah mengubah *term* yang berimbuhan menjadi *term* dengan kata dasar. Misalkan kata "bermain" menjadi "main" jika kata dalam Bahasa Indonesia.

2.3 Term Weighting

Term weighting atau pembobotan merupakan proses terakhir dari *preprocessing*. *Term weighting* adalah pemberian suatu nilai atau bobot pada *term*. Bobot yang dimiliki suatu term akan menjadi penentu klasifikasi suatu teks, maka proses ini dilakukan sebelum klasifikasi suatu teks. Pada tugas akhir ini akan digunakan metode pembobotan TF-IDF untuk mencari tingkat kepentingan kata sebagai vector ciri. Ada beberapa metode *term wighting* yaitu,

1. Term Frequency (TF)

Term Frequency (TF) merupakan metode pembobotan yang paling sederhana yaitu hanya dengan memberikan bobot untuk *term* yang sering pada sebuah dokumen. Pada metode ini, Setiap *term* yang lebih sering muncul pada sebuah dokumen akan dianggap lebih penting. Persamaannya adalah [2].

$$W(d, t) = TF(d, t) \quad (2.1)$$

Dimana $TF(d, t)$ adalah frekuensi kemunculan *term* t pada dokumen d.

2. Term Frequency – Inverse Document Frequency (TF – IDF)

Term Frequency – Inverse Document Frequency (TF – IDF) Adalah metode pembobotan yang semakin sedikit suatu kata muncul maka semakin tinggi bobotnya[8][9].

$$idf = \log\left(\frac{n}{df}\right) \quad (2.2)$$

n = jumlah teks

df = jumlah teks yang mengandung satu kata

2.4 Decision Tree C4.5

Klasifikasi merupakan suatu fungsionalitas yang akan menghasilkan model yang mampu memprediksi kelas atau kategori dari objek-objek [11]. *Decision tree* C4.5 merupakan salah satu metode klasifikasi. *Decision Tree* adalah teknik data mining untuk menyelesaikan masalah klasifikasi dengan model prediksi berupa pohon yang memiliki batang, daun, dan akar.. Jenis *decision tree* yang digunakan adalah *decision tree* C4.5 karena karakteristik data yang digunakan bersifat kontinu. C4.5 merupakan ekstensi dari algoritma *decision tree* ID3 (*iterative Dychotomizer version 3*) [3]. Kelebihan dari C4.5 dibanding ID3 adalah C4.5 dapat menangani atribut kontinu, dapat menangani *missing value*, dan memiliki proses runing untuk menghindari *overfit* (kondisi dimana klasifikasi berhasil pada data latih akan tetapi akan menurunkan nilai akurasi pada data uji) pada data latih.

Proses yang perlu dilakukan sebelum mengekstrak data ke dalam model *tree* adalah menentukan atribut yang akan menjadi root berdasarkan nilai gain ratio terbesar, kemudian menentukan atribut yang akan menjadi internal node untuk setiap branch dari node parent, dan membuat keputusan node ketika seleksi atribut tidak dapat digunakan. Persamaan yang digunakan dalam algoritma ini antara lain [4]:

1. Entropy

Entropi adalah suatu parameter untuk mengukur heterogenitas dari suatu data. Semakin kecil nilai Entropy maka semakin baik untuk digunakan dalam mengekstraksi suatu kelas.

$$Entropy(S) = \sum_i^c -p_i \log_2 p_i \quad (2.2)$$

Dimana:

p_i = proporsi sampel dalam kelas i

C = nilai dalam atribut target (angka kelas)

2. Information Gain

Nilai yang didapat dari perhitungan entropi masih belum asli tetapi, pengukuran efektivitas atribut dalam mengklasifikasikan data latih dapat ditentukan dengan informasi yang telah diperoleh. Dengan persamaan.

$$Gain(S, A) = Entropy(S) - \sum_{values(A)} \frac{|S_v|}{S} entropy(S_v) \quad (2.3)$$

Dimana:

S = nilai dari semua sampel data

A = atribut

Entropy (S) = entropi untuk sampel dengan nilai v

values (A) = koleksi nilai kemungkinan untuk atribut A

v = nilai kemungkinan untuk atribut A

$|S_v|$ = nilai sampel untuk nilai v

3. Gain Ratio

Gain Ratio merupakan modifikasi dari *information gain* untuk mengurangi bias atribut yang memiliki banyak cabang. Dimana persamaan ditulis sebagai.

$$Gain\ ratio = \frac{Gain(S,A)}{Split\ Information\ (S,A)} \tag{2.4}$$

Cari *Split Information* terlebih dahulu dengan cara.

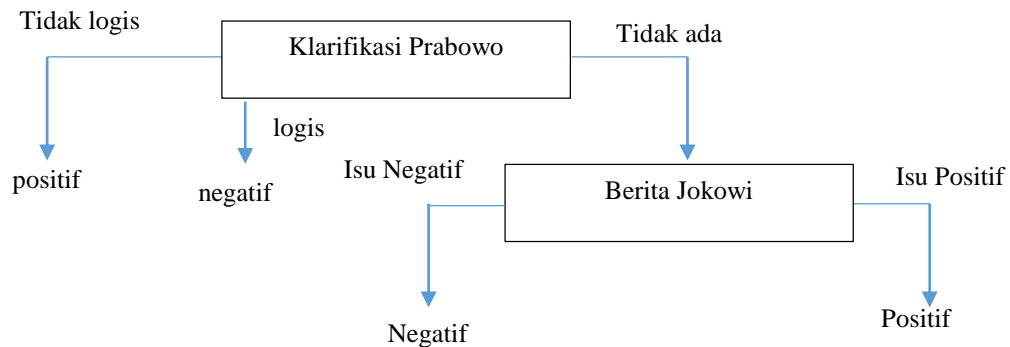
$$Split\ Information = - \sum_{t=1}^c \frac{s_i}{s} \log_2 \frac{s_i}{s} \tag{2.5}$$

Dimana nilai S_1 sampai S_c adalah subset dibuat dari splitting S menggunakan atribut A dengan varian C.

Table 2-1 Ilustrasi tabel *decision tree*

No.	Berita Jokowi	Bukti Jokowi	Bukti Prabowo	Klarifikasi Jokowi	Klarifikasi Prabowo	Hasil Sentimen terhadap Jokowi
1	Isu Positif	ada	ada	logis	tidak ada	positif
2	Isu Positif	ada	tidak ada	tidak ada	tidak ada	positif
3	Isu Positif	tidak ada	tidak ada	tidak ada	logis	negatif
4	Isu Negatif	tidak ada	ada	logis	logis	negatif
5	Isu Negatif	ada	ada	logis	tidak logis	positif
6	Isu Positif	tidak ada	tidak ada	tidak logis	logis	negatif
7	Isu Negatif	ada	ada	tidak ada	tidak ada	negatif

Tabel 2-1 adalah tabel ilustrasi tentang sentimen terhadap jokowi. Dapat diperhatikan bahwa banyak factor yang mempengaruhi berita tersebut. Seperti, jika Klarifikasi Prabowo tersebut logis, maka sentimen terhadap Jokowi itu negatif. Hasil mencari gain dan entropy dapat dilihat pada tabel 2-2 pada lampiran. Berikut hasil Pohon ilustrasi yang akan dihasilkan dari tabel 2-1:



Gambar 1 Ilustrasi *decision tree* untuk prediksi akhir untuk Berita Jokowi

2.5 Evaluasi Performasi

Performansi yang didapat dari setiap metode *term weighting* akan diukur menggunakan nilai akurasi, *precision*, dan *recall* menggunakan *confusion matrix*. Tabel 2-2 merupakan *confusion matrix*,

Table 2-2 tabel *confusion matrix*

Kelas Sebenarnya	Kelas Prediksi	
	Positif	Negatif
Positif	<i>True Positive</i>	<i>False Negative</i>
Negatif	<i>False Positive</i>	<i>True Negative</i>

Dimana:

True Positive (TP) = Kelas yang diprediksi benar, dan faktanya adalah benar.

True Negative (TN) = Kelas yang diprediksi salah, dan faktanya salah.

False Positive (FP) = Kelas yang diprediksi benar, dan faktanya salah.

False Negative (FN) = Kelas yang diprediksi salah, dan faktanya benar.

1. Akurasi

Tingkat kedekatan antara nilai prediksi dengan nilai sesungguhnya didefinisikan sebagai akurasi. Semakin besar nilai akurasinya, maka performansi klasifikasi semakin baik. Berikut persamaannya [5].

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} \quad (2.10)$$

2. Precision

Precision adalah tingkat ketepatan antara informasi yang telah dilakukan pengukuran ulang dengan nilai yang ditetapkan. Dalam *data mining*, *precision* adalah jumlah dokumen yang dengan benar diklasifikasikan dalam sebuah kelas dibagi jumlah total dokumen dalam kelas tersebut [6]. Dengan persamaan [5].

$$Precision(P) = \frac{TP}{TP + FP} \quad (2.11)$$

3. Recall

Recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Dalam *data mining*, *recall* dapat didefinisikan sebagai jumlah dokumen yang dengan benar diklasifikasikan dalam sebuah kelas dibagi jumlah total dokumen yang diklasifikasikan dalam kelas tersebut [6]. Dengan persamaan [5].

$$Recall(R) = \frac{TP}{TP + FN} \quad (2.12)$$

4. F - Measure

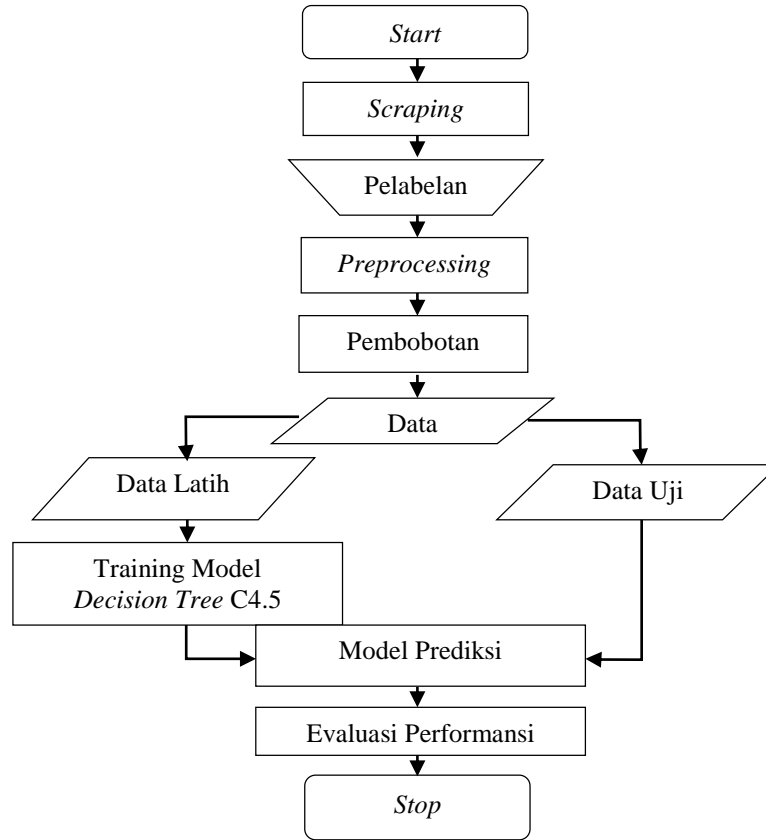
F-Measure digunakan untuk melakukan evaluasi nilai rata-rata Precision dan Recall. Semakin besar F-Measure, maka performa klasifikasi semakin tinggi.

$$F - Measure = \frac{2RP}{R + P} \quad (2.12)$$

3. Rancangan Sistem

Pada tugas akhir ini, sistem yang akan dibuat adalah sistem yang mengimplementasikan metode pembobotan TF dan TF-IDF dengan klasifikasi decision tree C4.5 untuk mengelompokkan teks berita yang memiliki sentimen positif atau negatif pada salah satu calon atau wakil calon presiden di pemilu 2019, diantaranya Jokowi-Ma'ruf dan Prabowo-Sandi. Data yang digunakan adalah teks berita yang didapatkan dari hasil *scrapping* yang sudah dilabelkan dari situs berita *online* Tempo. Hasil yang diharapkan bisa mengklasifikasikan positif atau negatifnya berita tersebut pada capres dan cawapres tersebut.

Rangkaian proses dalam sistem ini akan dibangun seperti gambar 2.



Gambar 2 Rancangan Sistem

4. Evaluasi

Evaluasi terdapat dua bagian, yaitu hasil uji dan analisis hasil uji.

4.1. Hasil Uji

Hasil uji berisi pelabelan dan klasifikasi *decision tree* C4.5

4.1.1. Pelabelan

Pelabelan dilakukan dengan mengelompokkan berita ke dalam kelas positif dan negatif pada kubu Prabowo – Sandi dan Jokowi – Ma’ruf.

Kriteria kelas negatif:

1. Berita membahas isu negatif salah satu paslon tanpa adanya pembelaan dari paslon yang disebarkan isu negatifnya.
2. Berita yang berisikan tentang paslon yang mengejek atau menuduh paslon lain tanpa adanya bukti yang kuat.
3. Berita yang berisi klarifikasi paslon tapi tidak logis dengan tidak adanya bukti.

Kriteria kelas positif:

1. Berita yang membanggakan atau menyebarkan isu isu positif paslon, misalnya seperti visi misi dan janji-janji yang menjanjikan.
2. Berita yang membantah isu negatif dengan jawaban yang logis dengan adanya bukti.

4.1.2. Klasifikasi Decision Tree C4.5

Tabel confusion matrix untuk Prabowo – Sandi ada pada tabel 4-1 sampai 4-5 dan untuk Jokowi – Ma’ruf ada pada tabel 4-6 sampai 4-10 pada lampiran. Pada tabel tersebut terdapat akurasi, *F-Measure*, *recall*, dan presisinya.

4.2 Analisis Hasil Uji

Berikut adalah tabel analisis hasil uji yang didapatkan dari klasifikasi *decision tree* C4.5 sebelumnya,

Tabel 4-1 Data Set dan Akurasi dari Prabowo – Sandi dan Jokowi – Ma’ruf menggunakan *Decision Tree* C4.5

Data Set	Akurasi	
	Prabowo – Sandi	Jokowi – Ma’ruf
0.9 : 0.1	93.33%	73.33%
0.8 : 0.2	83.33%	71.66%
0.7 : 0.3	81.32%	71.43%
0.6 : 0.4	81.66%	72.50%
0.5 : 0.5	82.00%	75.33%

Pada hasil percobaan di tabel 4-11 dapat dilihat bahwa total akurasi terbesar pada Prabowo-Sandi dan Jokowi – Ma'ruf ada pada data set 0.9 : 0.1.

Dapat dilihat dari gambar 3 yang ada pada lampiran bahwa sentiment positif pada kubu Jokowi (42%) lebih sedikit dibandingkan kubu Prabowo (43%).

Jika melihat tabel 5 pada lampiran, dapat diketahui bahwa rata-rata akurasi klasifikasi *Decision Tree* C4.5 lebih baik dibandingkan Naïve Bayes pada tabel 5 tersebut.

5. Kesimpulan

Pada tugas akhir ini dapat disimpulkan bahwa metode *decision tree* C4.5 dapat menentukan sebuah sentimen dengan kelas positif dan negatif pada sebuah berita. Pengklasifikasian menggunakan *decision tree* C4.5 dapat terbilang cukup baik karena akurasi pada data set 0.9 : 0.1 kubu Prabowo mendapatkan presentase akurasi 90% dan pada kubu Jokowi mendapatkan 73.33%. Untuk kedepannya riset ini akan dikembangkan lagi menjadi lebih seimbang.

Daftar Pustaka

- [1] B. Pang, L. Lee dan S. Vaithyanathan, "Thumbs Up? Sentiment Classification using Machine Learning," 2002.
- [2] T., Tokunaga. dan M., Iwayama. 1994. *Text Categorization Based On Weighted Inverse Document Frequency*. Tokyo, Japan: Tokyo Institute of Technology.
- [3] Sunjana. *Aplikasi Mining Data Mahasiswa dengan Metode Klasifikasi Decision Tree*. Universitas Widyatama.
- [4] Suyanto, ST, MSC.2007. *ARTIFICIAL INTELLIGENCE*. Bandung: Penerbit Informatika.
- [5] Guilet, Fabrice., Hamilton, Howard J. 2007. *Quality Measures in Data Mining*. Springer.
- [6] Golda Rossellini, Rebecca. 2012. *Perbandingan Metode Pembobotan Term Menggunakan Term Frequency Chi-Square dan Term Frequency Inverse Document Frequency pada Text Mining*. Bandung: IT Telkom.
- [7] Asrofi Buntoro, Ghulam. 2017. *Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter*. Ponorogo, Indonesia: Universitas Muhammadiyah Ponorogo.
- [8] S. Menaka dan N. Radha, "Text Classification using Keyword Extraction Technique," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 12, 2013.
- [9] S. Charanyaa dan K. Sangeetha, "Term Frequency Based Sequence Generation Algorithm for Graph Based Data Anonymization," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, no. 2, 2014.
- [10] Mursalat Ismail, Mehdi. 2019. *Sentimen Analisis pada Media Online Mengenai Pemilihan Presiden 2019 Menggunakan Metode Naïve Bayes*. Bandung, Indonesia: Telkom University.
- [11] Aradea, Satriyo A. dan Ariyan Z., Yuliana A, "Penerapan Decision Tree untuk Penentuan Pola Data Penerimaan Mahasiswa Baru". Semarang, Indonesia: Universitas Diponegoro, 2011.

Lampiran

Tabel 1 Pencarian Entropy dan Gain

node			Jumlah Kasus	Positif	Negatif	Entropy	Gain
1	total		7	3	4	0.9852	
	Berita Jokowi						0.0203
		Isu Positif	4	2	2	1	
		Isu Negatif	3	1	2	0.9182	
	Bukti Jokowi						0.5216
		Ada Bukti	4	3	1	0.8113	
		Tidak Ada Bukti	3	0	3	0	
	Bukti Prabowo						0.0203
		Ada Bukti	4	2	2	1	
		Tidak Ada Bukti	3	1	2	0.9182	
	Klarifikasi Jokowi						0.5666
		logis	3	2	1	0.9182	
		tidak logis	1	0	1	0	
		tidak ada	3	1	2	1	
	Klarifikasi Prabowo						0.5917
		logis	3	0	3	0	
	tidak logis	1	1	0	0		
	tidak ada	3	2	1	0.9182		

Tabel 2 Confusion Matrix Prabowo – Sandi dengan Data Set 0.9 : 0.1

Data Set 0.9 : 0.1		Prediction Class	
		Positive	Negative
Actual Class	Positive	25	0
	Negative	2	3

Accuracy	F-Measure	Precision	Recall
93.33%	0.96	0.93	1

Tabel 3 Confusion Matrix Prabowo – Sandi dengan Data Set 0.8 : 0.2

Data Set 0.8 : 0.2		Prediction Class	
		Positive	Negative
Actual Class	Positive	48	2
	Negative	8	2

Accuracy	F-Measure	Precision	Recall
83.33%	0.91	0.86	0.96

Tabel 4 Confusion Matrix Prabowo – Sandi dengan Data Set 0.7 : 0.3

Data Set 0.7 : 0.3		Prediction Class	
		Positive	Negative
Actual Class	Positive	68	4
	Negative	13	6

Accuracy	F-Measure	Precision	Recall
81.32%	0.89	0.84	0.94

Tabel 5 Confusion Matrix Prabowo – Sandi dengan Data Set 0.6 : 0.4

Data Set 0.6 : 0.4		Prediction Class	
		Positive	Negative
Actual Class	Positive	96	2
	Negative	20	2

Accuracy	F-Measure	Precision	Recall
81.66%	0.90	0.83	0.98

Tabel 6 Confusion Matrix Prabowo – Sandi dengan Data Set 0.5 : 0.5

Data Set 0.5 : 0.5		Prediction Class	
		Positive	Negative
Actual Class	Positive	118	8
	Negative	19	5

Accuracy	F-Measure	Precision	Recall
82%	0.90	0.86	0.94

Tabel 7 Confusion Matrix Jokowi – Ma’ruf dengan Data Set 0.9 : 0.1

Data Set 0.9 : 0.1		Prediction Class	
		Positive	Negative
Actual Class	Positive	21	5
	Negative	3	1

Accuracy	F-Measure	Precision	Recall
73.33%	0.84	0.87	0.81

Tabel 8 Confusion Matrix Jokowi – Ma’ruf dengan Data Set 0.8 : 0.2

Data Set 0.8 : 0.2		Prediction Class	
		Positive	Negative
Actual Class	Positive	41	7
	Negative	10	2

Accuracy	F-Measure	Precision	Recall
71.66%	0.83	0.80	0.85

Tabel 9 Confusion Matrix Jokowi – Ma’ruf dengan Data Set 0.7 : 0.3

Data Set 0.7 : 0.3		Prediction Class	
		Positive	Negative
Actual Class	Positive	59	14
	Negative	12	6

Accuracy	F-Measure	Precision	Recall
71.43%	0.82	0.83	0.81

Tabel 10 Confusion Matrix Jokowi – Ma’ruf dengan Data Set 0.6 : 0.4

Data Set 0.6 : 0.4		Prediction Class	
		Positive	Negative
Actual Class	Positive	77	20
	Negative	13	10

Accuracy	F-Measure	Precision	Recall
72.5%	0.82	0.86	0.79

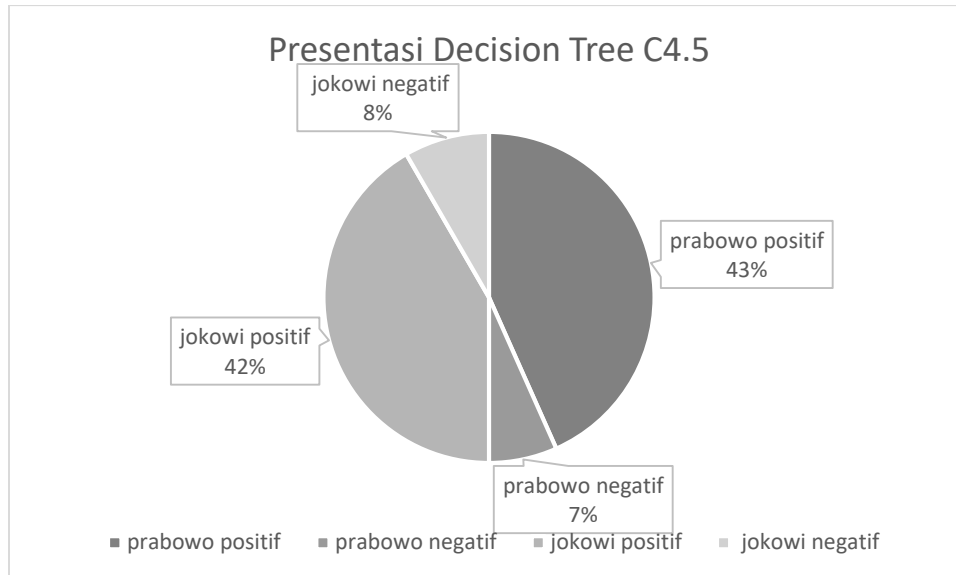
Tabel 11 Confusion Matrix Jokowi – Ma’ruf dengan Data Set 0.5 : 0.5

Data Set 0.5 : 0.5		Prediction Class	
		Positive	Negative
Actual Class	Positive	102	16
	Negative	21	11

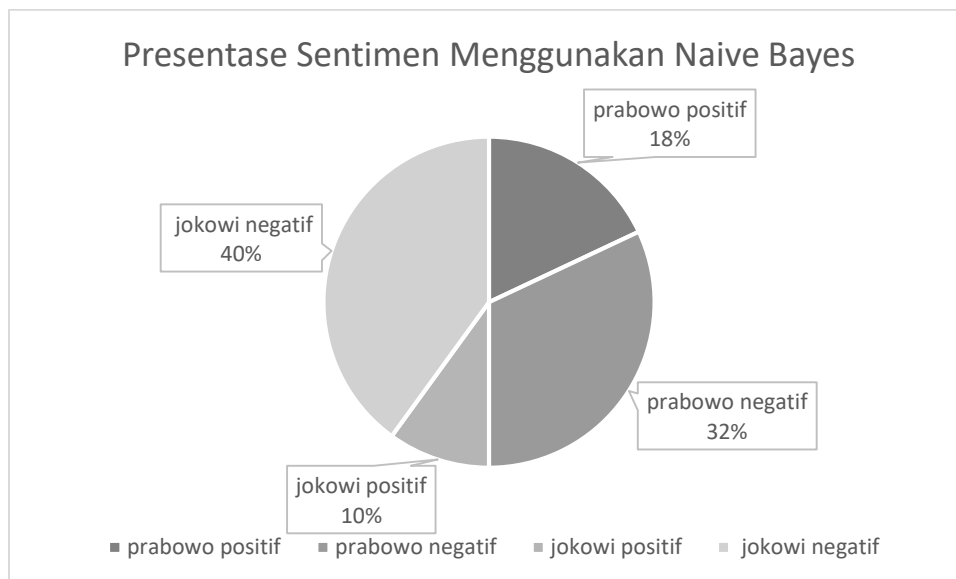
Accuracy	F-Measure	Precision	Recall
75.33%	0.85	0.83	0.86

Tabel 12 data set dan akurasi dari Prabowo – Sandi dan Jokowi – Ma’ruf menggunakan Naïve Bayes [10]

Jokowi-Ma’ruf		Prabowo-Sandi	
Rasio	Akurasi(%)	Rasio	Akurasi(%)
0.9 : 0.1	70	0.9 : 0.1	55
0.8 : 0.2	76.25	0.8 : 0.2	61.25
0.7 : 0.3	78.5124	0.7 : 0.3	57.8512
0.6 : 0.4	78.125	0.6 : 0.4	60
0.5 : 0.5	79.5	0.5 : 0.5	64



Gambar 1 presentase sentiment berita dengan *decision tree* C4.5



Gambar 2 presentase sentiment berita dengan Naïve Bayes [10]

```
Anaconda Prompt
      1      0.84      0.79      0.82      73
avg / total      0.74      0.71      0.72      91
Rasio data latih dan data test= 0.6 0.4
Accuracy: 71.66666666666667%
[[ 9 14]
 [20 77]]
      precision      recall      f1-score      support
-1      0.31      0.39      0.35      23
 1      0.85      0.79      0.82      97
avg / total      0.74      0.72      0.73      120
Rasio data latih dan data test= 0.5 0.5
Accuracy: 76.0%
[[19 13]
 [23 95]]
      precision      recall      f1-score      support
-1      0.45      0.59      0.51      32
 1      0.88      0.81      0.84      118
avg / total      0.79      0.76      0.77      150
```

Gambar 3 Screenshot hasil *runing* program sentimen *decision tree* C4.5