

Analisis Sentimen terhadap Pemerintahan Ridwan Kamil sebagai Gubernur Jawa Barat Menggunakan Algoritma Naïve Bayes

U. Darmanto Soer¹⁾, Sifa Fauziah²⁾, Sutrisno³⁾

¹⁾²⁾³⁾ Program Studi Teknik Informatika Fakultas Teknik, Universitas Pelita Bangsa

^{*)}Correspondence Author: darmantosoer@pelitabangsa.ac.id, Bekasi, Indonesia

DOI: <https://doi.org/10.37012/jtik.v9i2.1976>

Abstrak

Penggunaan sosial media di era globalisasi sangat diperlukan bagi sebagian kalangan tidak terkecuali pemimpin daerah. Dalam satu tahun masa jabatannya, Ridwan Kamil mendapatkan berbagai pujian maupun kritikan. Penelitian dilakukan untuk menganalisa sentimen masyarakat terhadap gubernur terpilih Ridwan Kamil. Pengumpulan data dilakukan dengan menggunakan proses *crawling* data Twitter menggunakan software Orange 3. Tahapan *preprocessing* terdiri dari proses *Remove Duplicates* yang bertujuan untuk memfilter data tweet yang sama, dan proses *Cleansing* yang bertujuan untuk membersihkan data dari *noise* atau gangguan. *Document Processing* terdiri dari beberapa proses berikut, *Transform Case*, *Tokenize*, *Filter Token by Length*, *Filter Stopwords*, *Stemming*, dan *Generate N-Grams*. Penelitian ini menggunakan algoritma Naïve Bayes untuk melakukan klasifikasi sentimen dan mencari nilai *preference value* dikarenakan algoritma tersebut memiliki akurasi yang cukup baik. Dari hasil pengujian yang dilakukan menggunakan teknik *cross validation* dan pengukuran akurasi menggunakan *confusion matrix* dengan dilakukan 10 kali pengujian akurasi terbaik yang diperoleh adalah 84,38 %. Respon positif masyarakat terhadap kepemimpinan Ridwan Kamil yang didapatkan dari hasil penghitungan *preference value* adalah 49%. Sedangkan nilai respon positif tersebut dapat berubah-ubah, dikarenakan respon masyarakat dan data yang diperoleh dapat berubah sewaktu-waktu. Dengan demikian dapat disimpulkan bahwa algoritma Naïve Bayes dapat digunakan untuk melakukan klasifikasi dengan cukup baik dan dapat mengukur respon masyarakat terhadap pemimpin daerah.

Kata kunci: Analisa Sentimen, Twitter, Naïve Bayes Classifier, Cross Validation, Preference Value

Abstract

The use of social media in the era of globalization is very necessary for some groups, including regional leaders. In his one year term of office, Ridwan Kamil received various praise and criticism. The research was conducted to analyze public sentiment towards the elected governor Ridwan Kamil. Data collection was carried out using the Twitter data crawling process using Orange 3 software. The preprocessing stage consists of the Remove Duplicates process which aims to filter the same tweet data, and the Cleansing process which aims to clean the data from noise or interference. Document Processing consists of the following processes, Transform Case, Tokenize, Filter Token by Length, Filter Stopwords, Stemming, and Generate N-Grams. This research uses the Naïve Bayes algorithm to carry out sentiment classification and find preference values because this algorithm has quite good accuracy. From the results of testing carried out using cross validation techniques and measuring accuracy using a confusion matrix by testing 10 times, the best accuracy obtained was 84.38%. The public's positive response to Ridwan Kamil's leadership obtained from the results of calculating preference value was 49%. Meanwhile, the positive response value can change, because the community's response and the data obtained can change at any time. Thus it can be concluded that the Naïve Bayes algorithm can be used to carry out classification quite well and can measure the public's response to regional leaders.

Keywords: Sentiment Analysis, Twitter, Naïve Bayes Classifier, Cross Validation, Preference Value

PENDAHULUAN

Perkembangan teknologi informasi yang semakin pesat di era globalisasi saat ini semakin memberikan kemudahan untuk berkomunikasi, contohnya adalah melalui internet. Menurut Kementerian Komunikasi dan Informatika (Kemenkominfo, 2013) pengguna internet di Indonesia saat ini mencapai 63 juta orang. Dari angka tersebut, terdapat 95 persennya menggunakan internet untuk mengakses media sosial. Salah satu media sosial yang diakses masyarakat adalah Twitter.

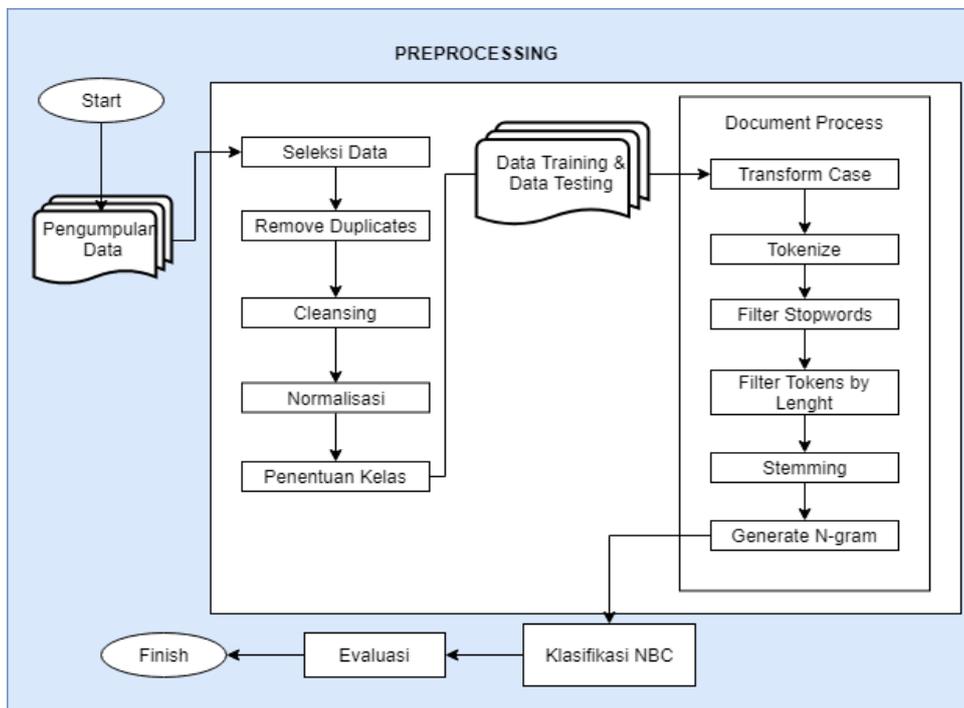
Twitter sebagai salah satu situs *microblogging* mengalami banyak perkembangan, terhitung pada 21 Maret 2016 Twitter genap memasuki usianya yang ke 10. Media sosial ini secara global memiliki 332 juta pengguna bulanan dengan 500 juta kicauan dikirim setiap hari dan 200 miliar kicauan dalam setahun (Maulana, 2016). Hal ini tentu saja menyebabkan semakin banyaknya pengguna diseluruh dunia yang berbagi dan membuat posting tentang produk dan layanan yang mereka gunakan atau mengekspresikan pandangan mereka tentang isu agama dan politik.

Twitter juga tidak jarang dimanfaatkan oleh beberapa orang untuk memudahkan pekerjaannya, sebagai contoh seorang gubernur yang menggunakan twitter sebagai media untuk mendengarkan keluhan-keluhan dan masukan dari warganya. Di Indonesia sendiri hal tersebut sudah dilakukan oleh beberapa kepala daerah yang aktif sekali bermain *mikro blogging* Twitter ini. Ridwan Kamil yang sampai saat ini telah memiliki 3,6 juta pengikut atau followers di akun Twitter miliknya yaitu @ridwankamil memanfaatkan Twitter dengan baik untuk menjaga kedekatan dengan warganya. Namun pemerintahannya sendiri tidak luput dari komentar publik dan menuai berbagai macam komentar. Dengan banyaknya pengguna Twitter yang menyampaikan opini-opini tersebut dapat dimanfaatkan untuk mencari informasi. Salah satu metode yang dapat digunakan untuk menganalisis opini-opini Twitter adalah analisa sentimen.

Penelitian ini dilakukan karena Ridwan Kamil sendiri mendapatkan banyak sekali dukungan dari masyarakat. Selain dukungan Ridwan Kamil sendiri mendapat berbagai macam kritik atau ujaran kebencian. Pandangan tersebut belum diklasifikasikan sehingga masyarakat belum tau berapa banyak tingkat sentimen positif dan negatif terhadap Ridwan Kamil.

METODE

Langkah-langkah penelitian sesuai dengan alur penelitian adalah sebagai berikut:



Gambar 1. Kerangka Berfikir Penelitian

Pengumpulan data dilakukan dengan menggunakan proses *crawling* data Twitter menggunakan software Orange 3. Data yang dikumpulkan merupakan tweet yang berbahasa Indonesia yang diambil dari akun twitter @ridwankamil yang berisi komentar para pengguna twitter.

title	Author	Content
1	@kompascom	BMKG sebelumnya mencatat sebanyak 72 kali gempa bermagnitudo di bawah 5 yang terja...
2	@inilahkorancom	Ridwan Kamil Dukung Pemanfaatan Teknologi untuk Bisnis di TTG IX/2019 https://t.co/nfZ...
3	@imuliii_	Cape anjir tadi cleaning area, gegara besok mau ada Ridwan Kamil ke tempat kerja:(
4	@Rcka_	@dimasdn Oh tidak, saya cinta ridwan kamil ❤️
5	@yaudehsi	Ridwan Kamil kalo jadi anak pertama jadinya
6	@Sky_marryland	@sukiono @CH_chotimah UAS itu sama aja dgn ustad Bikini yg dulu ribut sama Ridwan Ka...
7	@coxarchipela...	Kemaren oknum tidak berpendidikan banyak bgt yg protes masalah Illuminati masjid ranc...
8	@tempofoto	Gaya Ridwan Kamil dan Istri Bersepeda Saat Bekerja https://t.co/WyAXTgJuDU
9	@sbobetsa78	Hehree gubernur miskin penghargaan @ridwankamil
10	@davidmillit1	@fahiraidrisc @kumparan marah si ridwan kamil ntar..lahan duit untuk jabar jadi dikit...ekon...
11	@rss_tester	Ridwan Kamil Dukung Pemanfaatan Teknologi untuk Bisnis di TTG IX/2019 https://t.co/gL8t...
12	@kumparan	@fahiraidrisc @aniesbaswedana Di sisi lain, menurut Gubernur Jabar @ridwankamil, untuk m...
13	@Pemkot_Suka...	Penerapan teknologi merupakan salah satu pendorong pembangunan desa di Jawa Barat, ...
14	@antmegapolit...	https://t.co/p9SU4LkFLE.
15	@antmegapolit...	https://t.co/p9SU4LkFLE
16	@Arcani17	Sigana mah kitu...Trus kmaha kang ridwan kamil..??

Gambar 2. Pengumpulan Data

Seleksi Data bertujuan untuk menyeleksi data tweet yang akan digunakan nantinya, dikarenakan data tweet yang didapatkan dari proses *crawling* menggunakan Orange 3 belum sepenuhnya bersih dan memiliki data yang tidak diperlukan (Hasanudin, 2018).

Proses *Remove Duplicates* bertujuan untuk memfilter data tweet yang sama, hal ini disebabkan karena twitter memiliki fitur retweet yang menyebabkan banyaknya tweet yang sama di retweet oleh pengguna lain dan menimbulkan teks berulang dengan isi yang sama.

Cleansing bertujuan untuk membersihkan data dari *noise*. Seperti karakter HTML, ikon emosi, hashtag, username, url dan email (Buntoro, 2019). Proses Normalisasi mengganti kata yang tidak baku. Kata tidak baku merupakan kata yang tidak terdapat pada Kamus Besar Bahasa Indonesia (KBBI).

Data yang sudah melalui proses pada tahapan *preprocessing* kemudian akan dilakukan penentuan kelas pada atributnya. Penentuan ini diberikan sesuai subjektifitas peneliti, pembagian atribut pada penelitian ini sendiri hanya akan dibagi menjadi positif dan negatif.

Document Processing terdiri dari beberapa proses berikut:

1. *Transform Case*

Tahapan *transform cases* sendiri dalam penelitian ini bertujuan untuk menyeragamkan bentuk huruf menjadi *lower case* karena pada komentar Twitter sendiri memiliki bentuk huruf yang beragam.

2. *Tokenize*

Tokenize merupakan sebuah proses untuk memisahkan atau pemisahan dokumen menjadi per kata atau beberapa kata dalam suatu token/unit kecil.

3. *Filter Token by Length*

Filter Tokens by Length merupakan proses penghapusan kata atau token berdasarkan panjang karakter yang ditentukan. Pada penelitian ini minimum panjang karakter yang akan digunakan adalah empat karakter pada setiap kata.

4. *Filter Stopwords*

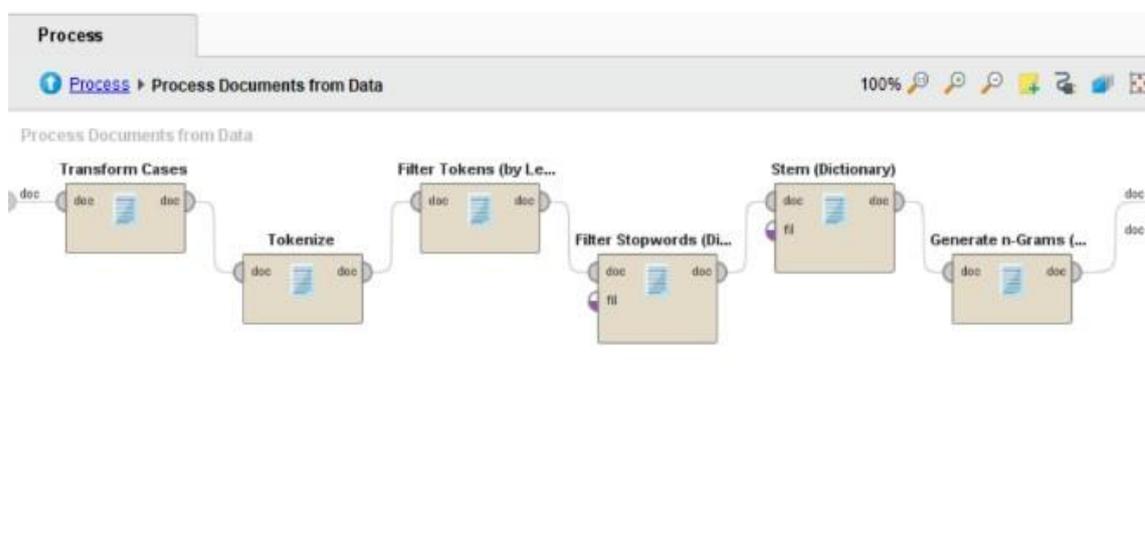
Tahapan *filter stopwords* (dictionary) merupakan tahapan proses yang bertujuan untuk menghapus atau mengfilter kata yang tidak memiliki arti dan tidak berhubungan dengan ucapan atau kata sentiment (Putra, 2017)

5. Stemming

Tahap Stemming adalah tahap menghilangkan imbuhan yang masih melekat sehingga dapat diperoleh kata dasar, dalam penelitian ini proses stemming akan dilakukan dengan menggunakan data *stem nolimit*.

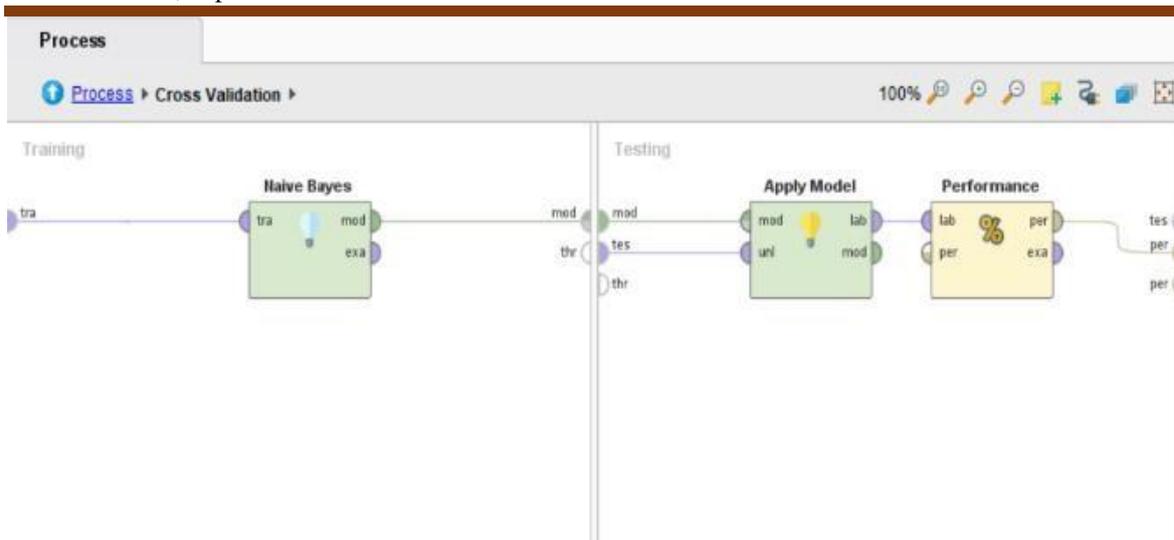
6. Generate N-Grams

Proses ini merupakan proses pemotongan string atau dokumen (n-gram) berdasarkan limit angka tertentu (Indhiarta, 2017). Dalam penelitian ini n-gram yang akan digunakan adalah bigram. Tujuan penggunaan N-gram sendiri karena dalam Bahasa Indonesia banyak frase yang tidak hanya terdiri dari satu kata.



Gambar 3. Document Processing

Dalam penelitian ini algoritma yang digunakan adalah Naïve Bayes Classifier dan berikut merupakan tampilan penerapan algoritma tersebut didalam sebuah sub proses pada operator *cross validation* dengan menambah kan Naïve Bayes pada tab training dan pada tab testing akan ditambahkan *apply model* dan *performance* yang nantinya akan digunakan sebagai penghitung performa klasifikasi.



Gambar 4. Klasifikasi *Naive Bayes Classifier (NBC)*

Dalam penelitian ini proses evaluasi akan menggunakan Confusion Matrix, Evaluasi bertujuan untuk mengetahui kinerja dari algoritma Naïve Bayes yang nantinya akan dijadikan sebagai acuan dari perbandingan akurasi pengujian (Turmudi, 2017).

Tabel 1. *Confusion Matrix*

		<i>Actual Class</i>	
		<i>Class 1</i>	<i>Class 2</i>
<i>Predicted</i>	<i>Class 1</i>	<i>True Positive</i>	<i>False Negative</i>
	<i>Class 2</i>	<i>False Positive</i>	<i>True Negative</i>

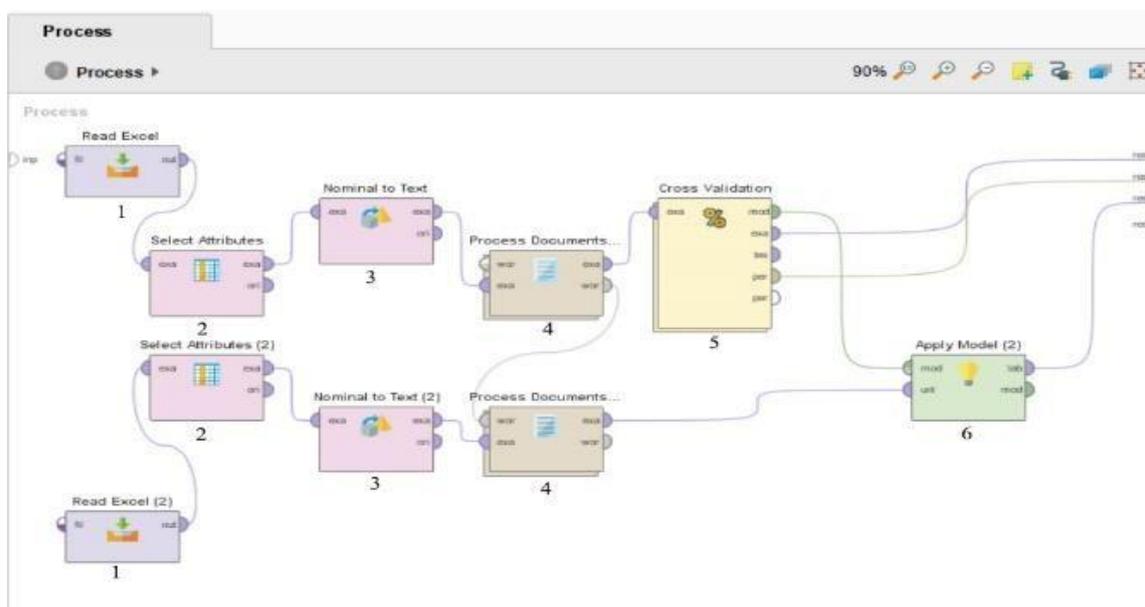
HASIL DAN PEMBAHASAN

Dataset yang digunakan dalam penelitian ini sebanyak ada 2500 tweet yang kemudian hanya dipilih sebanyak 1000 tweet yang masing-masing diambil dari pengguna yang mention ke akun @ridwankamil. Data diambil dari tanggal yang berbeda-beda dikarenakan twitter sendiri membatasi pengambilan data atau crawling selama 7 hari dari waktu pengambilan data. Untuk lebih jelasnya dapat dilihat pada tabel berikut:

Tabel 2. Jumlah Dataset

Dataset	Data Training	Data Testing	Total
@ridwankamil	800	200	1000

Proses Pengujian dilakukan dengan mengikuti alur yang telah ditentukan yaitu dengan metode klasifikasi Naïve Bayes dengan *Cross Validation*. Proses ini terdiri dari 10 kali proses pengujian yang mana nantinya akan dijadikan tolak ukur dari penelitian ini. Dalam proses ini peneliti menggunakan RapidMiner untuk melakukan pengujian dengan merangkai beberapa proses yaitu *read excel*, *Select Attribut*, *Nominal to text*, *Process Document*, *Cross Validation* dan *Apply Model*. Proses tersebut dapat dilihat pada gambar dibawah ini:



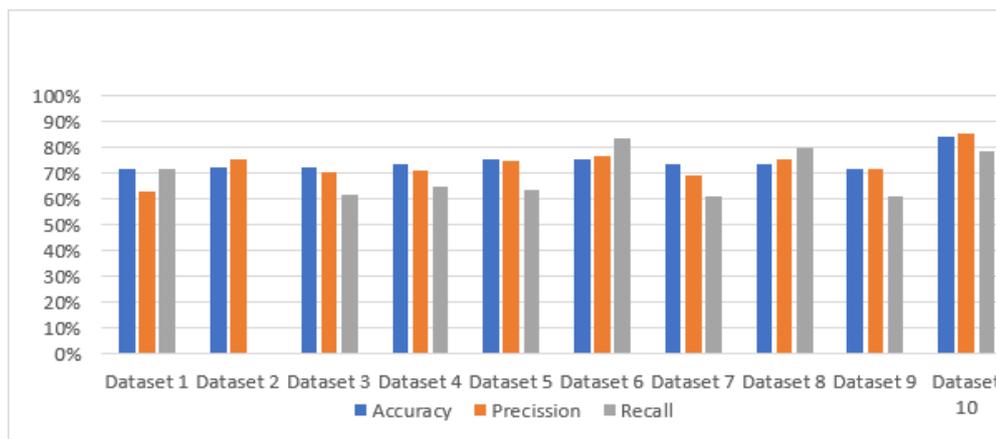
Gambar 5. Proses Pengujian

Dari hasil pengujian yang telah dilakukan menggunakan algoritma *Naïve Bayes* dan dilakukannya pengujian menggunakan teknik *cross validation* peneliti mendapatkan hasil pengujian seperti pada tabel berikut:

Tabel 3. Analisa Akurasi

Dataset	Accuracy	Precision	Recall
<u>Dataset 1</u>	71.62%	62.58%	71.69%
<u>Dataset 2</u>	72.35%	75.14%	78.28%
<u>Dataset 3</u>	72.35%	70.53%	61.78%
<u>Dataset 4</u>	73.59%	70.93%	65.00%
<u>Dataset 5</u>	75.59%	74.67%	63.36%
<u>Dataset 6</u>	75.50%	76.93%	83.32%
<u>Dataset 7</u>	73.22%	68.90%	61.13%
<u>Dataset 8</u>	73.47%	75.07%	79.83%
<u>Dataset 9</u>	71.50%	71.31%	61.14%
<u>Dataset 10</u>	84.38%	85.62%	78.75%

Dari tabel diatas maka dapat disimpulkan bahwa nilai *Accuracy*, *Precision* dan *recall* tertinggi terdapat pada pengujian ke 10 atau dataset 10. Pada dataset ini diperoleh hasil paling tinggi dengan nilai *Accuracy* 84.38% *Precision* 85.62% dan *Recall* sebanyak 78.75%. Hal ini menunjukkan bahwa penggunaan teknik *cross validation* cukup berpengaruh pada hasil pengujian yang didapatkan.



Gambar 6. Diagram Akurasi

Preference Value bertujuan untuk mengetahui respon positif masyarakat dari data yang telah diambil dan diproses sebelumnya. Jumlah data positif pada data training dan testing berjumlah 583 dan jumlah data negatif pada data training dan testing berjumlah 417. Total dari data tersebut berjumlah 1000 data, dan berikut adalah perhitungan *preference value* dari data tersebut dikalikan dengan nilai akurasi terbaik:

$$\textit{Preference Value} \frac{583}{583+417} * 84.38\% = 0,4919354$$

Semakin tinggi nilai *preference value* yang dihasilkan maka dapat diartikan semakin tinggi pula respon positif yang didapatkan (Ginting, Lhaksana, & Murdiansyah, 2018). Dan dari perhitungan diatas maka dapat disimpulkan bahwa respon positif masyarakat terhadap kepemimpinan Ridwan Kamil berdasarkan data yang diambil dari Twitter adalah 49%.

KESIMPULAN DAN REKOMENDASI

Berdasarkan penelitian yang telah dilakukan maka dapat ditarik kesimpulan bahwa penggunaan algoritma Naïve Bayes disertai dengan penggunaan teknik *Cross Validation* menghasilkan hasil yang cukup baik. Dengan nilai akurasi sebesar 84.38% dan tingkat respon positif masyarakat terhadap kepemimpinan Ridwan Kamil sebesar 49%. Sedangkan nilai respon positif tersebut dapat berubah-ubah, dikarenakan respon masyarakat dan data yang diperoleh dapat berubah sewaktu-waktu seperti data yang sekarang digunakan.

REFERENSI

- Buntoro, G. A. (2019). Analisis Sentimen Calon Gubernur Jawa Timur 2018 di Twitter. *Sci. Comput. Sci Informatics Journal*, 1(2), 37.
- Darwis, D, Siskawati, N, & Abidin, Z (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak*, ejurnal.teknokrat.ac.id,
<https://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/view/744>
- Ferdiana, R, Jatmiko, F, Purwanti, DD, & ... (2019). Dataset Indonesia untuk Analisis Sentimen. ... Teknik Elektro dan ..., journal.ugm.ac.id,
<https://journal.ugm.ac.id/v3/JNTETI/article/view/2558>
- Fikri, MI, Sabrila, TS, & Azhar, Y (2020). Perbandingan metode naïve bayes dan support vector machine pada analisis sentimen twitter. *SMATIKA Jurnal: STIKI Informatika Jurnal*, file.stiki.ac.id, <https://file.stiki.ac.id/SMATIKA/article/view/455>
- Ginting, H. S., Lhaksana, K. M., & Murdiansyah, D. T. (2018). Klasifikasi Sentimen Terhadap Bakal Calon Gubernur Jawa Barat 2018 di Twitter menggunakan Naive Bayes. *e-Proceeding Eng. 5*, hal. 1793-1802. Universitas Telkom.
- Hasanudin. (2018). Analisa sentimen terhadap tokoh publik pada twitter menggunakan metode klasifikasi naïve bayes classifier, term frequency, tf-idf dan n-gram.
- Indhiarta, W. C. (2017). *Penggunaan N-Gram pada analisa sentimen*.
- Maulana, A. (2016). *Twitter Rahasiakan Jumlah Pengguna di Indonesia*. Diambil kembali dari <https://www.cnnindonesia.com/teknologi/20160322085045-185-118939/twitter-rahasiakan-jumlah-pengguna-di-Indonesia>.
- Putra, R. S. (2017). *Analisis sentimen twitter dengan klasifikasi naïve bayes menggunakan seleksi fitur mutual information dan inverse document frequency*.
- Samsir, S, Ambiyar, A, Verawardina, U, & ... (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes. *Jurnal Media ...*, stmik-budidarma.ac.id, <http://stmik-budidarma.ac.id/ejurnal/index.php/mib/article/view/2580>

-
- Sari, FV, & Wibowo, A (2019). Analisis sentimen pelanggan toko online Jd. Id menggunakan metode Naïve Bayes Classifier berbasis konversi ikon emosi. Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu ..., jurnal.umk.ac.id, <https://jurnal.umk.ac.id/index.php/simet/article/view/3487>
- Singgale, YA (2021). Pemilihan metode dan algoritma dalam analisis sentimen di media sosial: systematic literature review. Journal of Information Systems and Informatics, journal-isi.org, <https://www.journal-isi.org/index.php/isi/article/view/125>
- Turmudi, A. (2017). Comparison Algorithm Classification Naive Bayes, Decision Tree, and Neural Network for Analysis Sentiment. Progr. Stud. Tek. Inform. Sekolah Tinggi Teknologi Pelita Bangsa, 12, 200.
- Watrianthos, R, Giatman, M, & ... (2022). Analisis Sentimen Pembelajaran Campuran Menggunakan Twitter Data. Jurnal Media ..., ejurnal.stmik-budidarma.ac.id, <http://www.ejurnal.stmik-budidarma.ac.id/index.php/mib/article/view/3383>
- Yulita, W (2021). Analisis sentimen terhadap opini masyarakat tentang vaksin covid-19 menggunakan algoritma naïve bayes classifier. Jurnal Data Mining dan Sistem Informasi, ejurnal.teknokrat.ac.id, <https://ejurnal.teknokrat.ac.id/index.php/JDMSI/article/view/1344>