

Analisis Sentimen Cuitan di Terhadap Kawasan Wisata Danau Toba Dengan Metode Naïve Bayes

Afryandani M Sidabutar¹, Indra M Sarkis², Edward Rajagukguk³
^{1,2,3}Fakultas Ilmu Komputer, Universitas Methodist Indonesia

Info Artikel

Histori Artikel:

Received, Maret, 2023
Revised, Maret, 2023
Accepted, April, 2023

Keywords:

Twitter,
Sentiment Analysis,
Naïve Bayes,
R-Studio,
Rapidminer

ABSTRACT

Teknologi yang semakin berkembang pesat memudahkan berbagai urusan dalam bentuk digital. Sosial media adalah salah satu yang memudahkan dalam berbagai aspek di dunia digital. Saat ini berbagai hal telah dibicarakan dalam berbagai platform sosial media, salah satunya adalah dalam sosial media Twitter. Sentimen analisis adalah salah satu contoh pemanfaatan sosial media yang banyak dilakukan saat ini. Sentimen analisis ini berguna untuk mengetahui pendapat berbagai macam orang dalam suatu topik atau isu. Danau toba sebagai tempat wisata yang sedang dipandang oleh berbagai kalangan sebagai tempat wisata prioritas menjadi salah satu yang paling banyak dibicarakan di media sosial. Oleh karena itu dibuat sentiment analisis terhadap Kawasan wisata Danau Toba dengan cara skoring atau pemberian nilai di tiap katanya. Naïve Bayes Classifier juga akan diterapkan dalam klasifikasinya. Klasifikasi Naïve Bayes akan bekerja berdasarkan pengalaman sebelumnya, dimana data tweet yang telah diberi label Positif, Negatif, dan Netral akan dilabeli Kembali dengan bantuan Aplikasi Rapidminer untuk. Akurasi dari sistem juga akan di uji dengan 10-Fold Cross Validation dan Naïve Bayes dengan aplikasi Rapidminer. Dari sistem Analisis sentimen menggunakan skoring yang telah dibuat, didapat hasil akurasi sebesar 74,57%, Recall 75,31% dan Precision 71,05%.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Penulis Koresponden:

Afryandani M Sidabutar
Universitas Methodist Indonesia, Medan,
Jl. Hang Tuah No.8, Medan - Sumatera Utara.
Email: alexsonndruru05@gmail.com

1. PENDAHULUAN

Pengguna Twitter kerap terlibat dalam topik yang sedang hangat dibicarakan di dunia maya. Tidak heran jika trending topic Twitter banyak didominasi oleh topik asal Indonesia. Twitter digunakan tidak hanya untuk berbagi status atau cerita, Twitter telah menjadi sarana promosi dan kampanye secara elektronik [1]. Dengan adanya Twitter, sebuah topik dapat menjadi sebuah fenomena yang mendunia apabila menarik para pengguna Twitter. Namun, informasi yang dihadirkan harus memiliki nilai yang penting dan menarik [2].

Analisis sentimen sering juga disebut sebagai opinion mining. Hal ini menunjukkan emosi apa yang ada di balik setiap kata-kata penulis. Data yang akan ada akan dilakukan proses analisis data

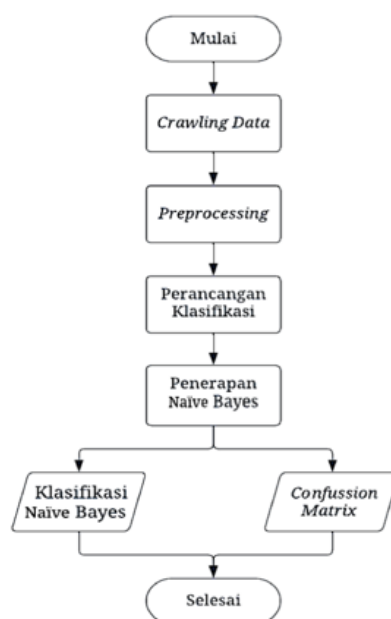
tulisan daring untuk menentukan nada emosional dari penulisnya. Data yang akan diolah untuk melakukan Analisis Sentimen akan di ambil dengan Web crawling. Data yang telah disiapkan akan di simpan pada Microsoft Excel dan akan dilakukan cleaning data untuk membersihkan emoji dan karakter yang tidak diperlukan. Untuk melakukan analisis sentiment terhadap data yang telah ada, penulis akan menggunakan metode Naïve Bayes Classifier. Naive Bayes classifier merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes.

Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yaitu memprediksi peluang berdasarkan pengalaman di masa sebelumnya (Teorema Bayes) dengan ciri utamanya adalah asumsi yang sangat kuat (naif) akan ketergantungan dari masing-masing kondisi/kejadian.

2. METODE PENELITIAN

2.1. Tahapan Perancangan

Gambar 1 merupakan tahapan perancangan dalam penelitian ini.



Gambar 1 Tahapan Perancangan

Perancangan dimulai dengan pengumpulan data Twitter secara scrapping atau crawling berdasarkan kata kunci “danau toba” melalui aplikasi R-Studio. Dilanjutkan dengan Preprocessing data tweet, setelah data menjadi bersih, kemudian data diklasifikasi menggunakan Sentimen Skor yang telah di tentukan dengan kumpulan kata Bahasa Indonesia. Naïve Bayes akan diterapkan untuk menghasilkan klasifikasi baru berdasarkan label klasifikasi yang telah dibuat. Confussion Matrix juga akan dihitung untuk mengevaluasi sistem yang telah dibuat.

2.2. Teknik Pengumpulan Data

Data yang digunakan pada penelitian ini diambil dari kumpulan tweets. Data tweets ini diperoleh dengan membuat program crawling yang menggunakan Tweet API. Dalam proses pengambilan data dibutuhkan Package TwitteR pada RStudio, twitteR (ditulis oleh Geoffjentry) adalah R package yang menyediakan akses ke API Twitter sehingga memungkinkan kita melakukan crawling data Twitter menggunakan R.

2.3. Preprocessing Data

Setelah data menjadi terstruktur dan mempunyai nilai numerik maka data bisa diproses untuk mengekstrak informasi dari dokumen-dokumen teks yang bisa digunakan untuk kepentingan

analisis berbagai bidang multidisipliner seperti klasifikasi. Adapun tahapan preprocessing data adalah sebagai berikut:

1. Cleansing, yaitu proses di mana menghapus atribut atau tanda baca yang tidak diperlukan pada proses analisis.
2. Case folding, yakni mengubah semua huruf besar menjadi lowercase atau huruf kecil.
3. Tokenizing, merupakan proses pemisahan kata berdasarkan tiap kata yang menyusunnya menjadi potongan tunggal.
4. Filtering, yaitu tahap menghapus kumpulan dari kata yang tidak berpengaruh tanpa mengurangi informasi dari dokumen. Dengan memanfaatkan stopwords yang telah disusun dan dilakukan penambahan kata secara manual.
5. Stemming merupakan proses mengubah kata menjadi bentuk dasarnya. Tujuan dari proses stemming adalah menghilangkan imbuhan-imbuhan.

2.4. Perancangan Klasifikasi

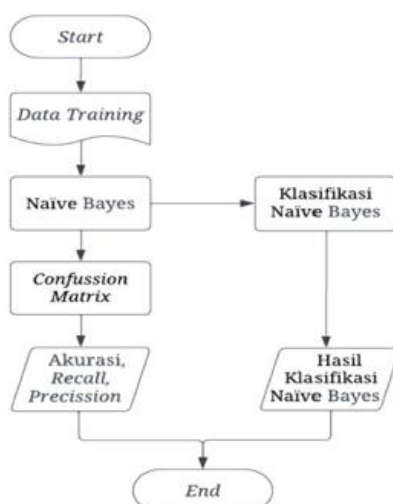
Perancangan klasifikasi dilakukan untuk mengelompokkan Tweet berdasarkan kelas yang ditentukan yang berisi nilai dari atribut-atribut yang ditentukan sesuai dengan parameter yang digunakan (jumlah kata positif, jumlah kata netral, jumlah kata negatif). Hasil klasifikasi adalah kandungan sentimen (positif, netral, negatif). Setiap kata yang memenuhi salah satu dari klasifikasi tersebut dikumpulkan dalam masing-masing variabel teks penampungan untuk diproses pada tahapan selanjutnya. Perancangan klasifikasi dapat dilihat pada Gambar 2.



Gambar 2 Perancangan Klasifikasi

2.5. Penerapan Naïve Bayes

Dari hasil klasifikasi, selanjutnya dilakukan pengujian dengan Naïve Bayes, berikut adalah alur proses dari Naïve Bayes. Dari 706 data yang di dapatkan pada tanggal 29 Juni 2022, akan di ambil data dengan perbandingan sebanyak 70:30. Di ambil 210 data dari 706 data yang akan digunakan untuk sebagai data testing. Data ini akan diberi label sesuai dengan prediksi sentiment yang sebelumnya telah dibuat. Data akan berisi 70 data positif, 70 data netral dan 70 data negatif. Data yang diberi label akan memprediksi klasifikasi baru dengan algoritma Naïve Bayes. hasil dari Naïve Bayes ini juga akan menunjukkan akurasi, recall, dan precision dari data yang sudah di klasifikasi sebelumnya. Tahapan Naïve Bayes dapat dilihat pada Gambar 3.



Gambar 3 Tahapan Naive Bayes

3. HASIL DAN PEMBAHASAN

3.1. Hasil Klasifikasi *Skoring*

Dari sentiment analisis yang dilakukan dengan memberi nilai/skor pada setiap kata telah didapatkan hasil sebagai berikut.

Negatif	Netral	Positif
199	411	96

Gambar 4 Hasil Klasifikasi

3.2. Hasil Naive Bayes

Data yang telah melalui tahap pemrosesan klasifikasi teks kemudian dapat dilanjutkan ke tahap selanjutnya yaitu klasifikasi dengan algoritma Naive Bayes Classifier. Data dalam bentuk teks akan muncul hasil klasifikasi teks yang mengandung positif, netral, dan negatif. Akan didapatkan juga akurasi dari sistem klasifikasi yang telah dibuat.

3.2.1. Hasil Klasifikasi Naive Bayes

Diambil 5 dokumen yang sebelumnya telah dilakukan proses sentimen analisis, dan telah di labeli dengan nilai sentimennya seperti pada Tabel 1.

Tabel 1 Data yang akan di proses

	Text	Nilai
D1	mari kenal destinasi wisata super prioritas	positif
D2	wisata pantai danau toba indah alami	positif
D3	danau toba salah lokasi syuting film ngeri ngeri sedap wisata ikonik sumatra utara	negatif
D4	asli cakep sangat danau toba	netral
D5	renang sampai danau toba	netral

Dari tabel di atas ditemukan 12 *term* positif, 14 *term* negatif, dan 9 *term* netral. Dari setiap *term* yang ditemukan akan dihitung probabilitas kemunculannya untuk setiap kelas dan dibagi total *term* yang ditemukan. *Term* setiap kelas yang ditemukan terdapat di Tabel 2.

Tabel 2 *Term* setiap kelas

kata	Positif	Negatif	Netral
alami	1/12	0/14	0/9
asli	0/12	0/14	1/9
cakep	0/12	0/14	1/9
danau	1/12	1/14	2/9
destinasi	1/12	0/14	0/9
ikonik	0/12	1/14	0/9
indah	1/12	0/14	0/9
kenal	1/12	0/14	0/9
lokasi	0/12	1/14	0/9
mari	1/12	0/14	0/9
ngeri	0/12	1/14	0/9
pantai	1/12	0/14	0/9
prioritas	1/12	0/14	0/9
renang	0/12	1/14	1/9
salah	0/12	1/14	0/9
sampai	0/12	1/14	1/9
sangat	0/12	0/14	1/9
satu	0/12	1/14	0/9
sedap	0/12	1/14	0/9
sumatera	0/12	1/14	0/9
super	1/12	0/14	0/9
syuting	0/12	1/14	0/9
toba	1/12	1/14	2/9
utara	0/12	1/14	0/9
wisata	2/12	1/14	0/9

Dalam proses perhitungan kemungkinan *term* tidak terdapat dalam suatu kelas sangat besar. Apabila *term* tersebut tidak terdapat suatu kelas maka *term* tersebut akan bernilai 0. Hal ini akan menyebabkan seluruh perhitungan Naïve Bayes menjadi tidak tepat. Terdapat sebuah metode bernama *Laplace Smoothing* yang biasa dipergunakan untuk mengantisipasi hal seperti ini.

Laplacian Smoothing merupakan metode *smoothing* yang dapat diterapkan pada Naïve Bayes Classifier. Konsepnya sangat sederhana yaitu dengan menambahkan nilai positif yang kecil pada setiap nilai probabilitas kondisional yang ada sehingga nilai yang tidak memiliki *term* dapat menghindari nilai nol pada model probabilitas dan mendapat nilai. Konsep ini diharapkan dapat mempermudah penerapan metode Naïve Bayes agar hasil yang didapatkan maksimal dan lebih akurat.

Jadi berdasarkan pernyataan di atas, kurangnya data pelatihan pada teori Naive Bayes dapat menyebabkan nilai probabilitas 0, cara menghindari atau mengatasinya dengan menggunakan metode *Laplace Smoothing* dengan menambahkan setiap perhitungan dengan angka 1. Penambahan ini tidak akan membuat perbedaan yang berarti, dan akan menghindari hasil nilai probabilitas 0. Maka di tambahkan nilai k yang mewakili *laplace smoothing* yang bernilai 1. Maka rumus Naïve Bayes akan diubah menjadi.

$$P(H|X) = \frac{P(X|H)P(H)+k}{P(X)}$$

Berikut adalah perhitungan untuk *term* “danau” pada kelas positif, Negatif dan netral.

1. Probabilitas kata kelas positif

$$P(\text{danau}|\text{positif}) = \frac{1+1}{12+35} = \frac{2}{47} = 0,04255319$$

2. Probabilitas kata kelas negatif

$$P(\text{danau}|\text{negatif}) = \frac{1+1}{14+27} = \frac{2}{49} = 0,04081633$$

3. Probabilitas kata kelas netral

$$P(\text{danau}|\text{netral}) = \frac{2+1}{9+35} = \frac{3}{44} = 0,06818182$$

Dari perhitungan di atas bisa di ambil nilai probabilitas dari semua kata dan dapat dilihat di Tabel 3.

Tabel 3 Nilai Setiap *Term*

kata	Positif	Negatif	Netral
alami	0,04255319	0,02040816	0,02272727
asli	0,0212766	0,02040816	0,04545455
cakep	0,0212766	0,02040816	0,04545455
danau	0,04255319	0,04081633	0,06818182
destinasi	0,04255319	0,02040816	0,02272727
ikonik	0,0212766	0,04081633	0,02272727
indah	0,04255319	0,02040816	0,02272727
kenal	0,04255319	0,02040816	0,02272727
lokasi	0,0212766	0,04081633	0,02272727
mari	0,04255319	0,02040816	0,02272727
ngeri	0,0212766	0,04081633	0,02272727
pantai	0,04255319	0,02040816	0,02272727
prioritas	0,04255319	0,02040816	0,02272727
renang	0,0212766	0,04081633	0,04545455
salah	0,0212766	0,04081633	0,02272727
sampai	0,0212766	0,04081633	0,04545455
sangat	0,0212766	0,02040816	0,04545455
satu	0,0212766	0,04081633	0,02272727
sedap	0,0212766	0,04081633	0,02272727
sumatera	0,0212766	0,04081633	0,02272727
super	0,04255319	0,02040816	0,02272727
syuting	0,0212766	0,04081633	0,02272727
toba	0,04255319	0,04081633	0,06818182
utara	0,0212766	0,04081633	0,02272727
wisata	0,06382979	0,04081633	0,02272727

Nilai probabilitas kata pada masing-masing kelas tersebut kemudian akan digunakan untuk menguji data yang ada, sedangkan probabilitas setiap kelas ditentukan dengan menggunakan rumus di bawah ini

$$P(c) = \frac{N_c}{N}$$

Keterangan:

N_c = Jumlah dokumen training dalam kelas c

N = Jumlah keseluruhan dokumen training dari seluruh kelas

Nilai prior Probabilities dapat dilihat dalam Tabel 4.

Tabel 4 Nilai Probabilitas Kelas

nilai	Probabilitas kelas
positif	$P(n)=\frac{2}{5}=0,4$
negatif	$P(n)=\frac{1}{5}=0,2$
netral	$P(n)=\frac{2}{5}=0,4$

Langkah pertama yang dilakukan untuk melakukan klasifikasi adalah memecah kalimat-kalimat dalam kata menjadi kata demi kata kemudian menghitung nilai probabilitas kata pada setiap kelas menggunakan tabel probabilitas kata dan probabilitas kelas yang telah diperoleh sebelumnya.

Hasil nilai probabilitas setiap kelas akan dijumlahkan dengan setiap nilai probabilitas setiap term yang di temukan dalam dokumen. Hasil nilai dari setiap kelas dengan nilai kelas tertinggi akan menjadi klasifikasi baru berdasarkan metode Naïve Bayes yang prinsipnya memanfaatkan pengalaman dari data sebelumnya digunakan untuk prediksi klasifikasi baru. Dari hasil di atas bisa kita ambil prediksi klasifikasi yang baru sebagai berikut:

1. Dokumen D1 bisa kita ambil sebagai berikut

Tabel 5 Nilai Probabilitas Dokumen D1

kelas	positif	negatif	netral
nilai kelas	0,4	0,2	0,4
term			
mari	0,04255319	0,02040816	0,02272727
kenal	0,04255319	0,02040816	0,02272727
destinasi	0,04255319	0,02040816	0,02272727
wisata	0,06382979	0,04081633	0,02272727
super	0,04255319	0,02040816	0,02272727
prioritas	0,04255319	0,02040816	0,02272727
nilai probabilitas	0,67659574	0,34285714	0,53636364
probabilitas tertinggi	Positif		

Nilai klasifikasi dari dokumen D1 dengan metode Naïve Bayes adalah Positif

2. Dokumen D2 bisa kita ambil sebagai berikut

Tabel 6 Nilai Probabilitas Dokumen D2

Kelas	positif	negatif	netral
nilai kelas	0,4	0,2	0,4
term			
wisata	0,06382979	0,04081633	0,02272727
pantai	0,04255319	0,02040816	0,02272727
danau	0,04255319	0,04081633	0,06818182
toba	0,04255319	0,04081633	0,06818182
indah	0,04255319	0,02040816	0,02272727
alami	0,04255319	0,02040816	0,02272727
nilai probabilitas	0,67659574	0,38367347	0,62727273
probabilitas tertinggi	positif		

Nilai klasifikasi dari dokumen D2 dengan metode Naïve Bayes adalah Positif.

3. Dokumen D3 bisa kita ambil sebagai berikut

Tabel 7 Nilai Probabilitas Dokumen D3

kelas	positif	negatif	netral
nilai kelas	0,4	0,2	0,4
term			
danau	0,04255319	0,04081633	0,06818182
toba	0,04255319	0,04081633	0,06818182
salah	0,0212766	0,04081633	0,02272727
satu	0,0212766	0,04081633	0,02272727
lokasi	0,0212766	0,04081633	0,02272727
syuting	0,0212766	0,04081633	0,02272727
ngeri	0,0212766	0,04081633	0,02272727
sedap	0,0212766	0,04081633	0,02272727
wisata	0,06382979	0,04081633	0,02272727
ikonik	0,0212766	0,04081633	0,02272727
sumatera	0,0212766	0,04081633	0,02272727
utara	0,0212766	0,04081633	0,02272727
nilai probabilitas	0,74042553	0,68979592	0,76363636
probabilitas tertinggi		netral	

Nilai klasifikasi dari dokumen D3 dengan metode Naïve Bayes adalah Netral.

4. Dokumen D4 bisa kita ambil sebagai berikut

Tabel 8 Nilai Probabilitas Dokumen D4

kelas	positif	negatif	netral
nilai kelas	0,4	0,2	0,4
term			
asli	0,0212766	0,02040816	0,04545455
cakep	0,0212766	0,02040816	0,04545455
sangat	0,0212766	0,02040816	0,04545455
danau	0,04255319	0,04081633	0,06818182
toba	0,04255319	0,04081633	0,06818182
nilai probabilitas	0,54893617	0,34285714	0,67272727
probabilitas tertinggi		netral	

Nilai klasifikasi dari dokumen D4 dengan metode Naïve Bayes adalah Netral.

5. Dokumen D5 bisa kita ambil sebagai berikut

Tabel 9 Nilai Probabilitas Dokumen D5

kelas	positif	negatif	netral
nilai kelas	0,4	0,2	0,4
term			
renang	0,0212766	0,04081633	0,04545455
sampai	0,0212766	0,04081633	0,04545455
danau	0,04255319	0,04081633	0,06818182
toba	0,04255319	0,04081633	0,06818182
nilai probabilitas	0,52765957	0,36326531	0,62727273

probabilitas tertinggi

netral

Nilai klasifikasi dari dokumen D5 dengan metode Naïve Bayes adalah Netral.

Kelima dokumen telah dilakukan proses klasifikasi dengan Naïve Bayes dan dari klasifikasi baru yang telah dibuat dengan metode Naïve Bayes di atas Maka di dapatkan hasil pada Tabel 10.

Tabel 10 Klasifikasi Baru

Text	Nilai	NBC
D1 mari kenal destinasi wisata super prioritas	positif	Positif
D2 wisata pantai danau toba indah alami	positif	Positif
D3 danau toba salah lokasi syuting film ngeri ngeri sedap wisata ikonik sumatra utara	negatif	Netral
D4 asli cakep sangat danau toba	netral	Netral
D5 renang sampai danau toba	netral	Netral

3.2.2. Confussion Matrix

Confusion matrix merupakan salah satu metode yang dapat digunakan untuk mengevaluasi kinerja suatu metode klasifikasi. Pada pengukuran kinerja menggunakan *confusion matrix*, terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). Nilai *True Negative* (TN) merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan *False Positive* (FP) merupakan data negatif namun terdeteksi sebagai data positif. Karena adanya nilai netral, maka di tambahkan nilai *True Netral* (TNet) dan *False Netral* (FNet). Akan dihitung nilai akurasi, nilai *recall* dan nilai *precision* dari sistem yang telah dibuat. Ketiga nilai ini bisa dihitung dari *confussion matrix* yang telah dibuat dengan model Rapidminer. Maka didapatkan *confussion matrix* seperti dibawah ini.

Tabel 11 *Confussion Matrix*

	true netral	true positif	true negatif
pred. netral	61	10	17
pred. positif	6	53	9
pred. negatif	3	7	44

1. Akurasi

Akurasi adalah nilai ketepatan suatu klasifikasi dalam bentuk persen, dimana akurasi dapat dihitung dengan persamaan berikut.

$$\text{Akurasi} = \frac{\text{TP} + \text{TNet} + \text{TN}}{\text{Jumlah Data}} \times 100$$

$$\text{Akurasi} = \frac{61 + 53 + 44}{210} \times 100 = 75,24\%$$

Dari hasil di atas maka di ketahui bahwa akurasi dari sistem yang dilakukan adalah 75,24%.

2. Recall

Recall digunakan untuk menghitung performansi dokumen tertentu, yang dapat dihitung dalam persamaan di bawah ini.

$$\text{Class Recall} = \frac{\text{Jumlah Data Benar}}{\text{Jumlah Data}} \times 100$$

$$\text{Class Recall Positif} = \frac{61}{70} \times 100 = 87,14\%$$

$$\text{Class Recall Netral} = \frac{53}{70} \times 100 = 75,71\%$$

$$\text{Class Recall Negatif} = \frac{44}{70} \times 100 = 62,86\%$$

Rata-Rata dari *Class recall* akan dihitung dari total *Class Recall* dan dibagi dengan jumlah kelas yang ada. Rata-rata recall dihitung sebagai berikut.

$$\text{Recall Avarage} = \frac{\text{Jumlah Recall}}{\text{Jumlah Kelas}}$$

$$\text{Recall Avarage} = \frac{87,14\% + 75,71\% + 62,86\%}{3} = 75,24\%$$

Dari hasil di atas maka di ketahui bahwa recall dari sistem yang dilakukan adalah 75,24%

3. Precision

Precision merupakan rasio prediksi benar dengan nilai kelasnya dibandingkan dengan keseluruhan hasil yang diprediksi dengan nilai kelasnya, *Precision* dihitung dengan rumus di bawah ini.

$$\text{Class Precision} = \frac{\text{Jumlah Data Benar}}{\text{Jumlah Data Kelas}} \times 100$$

$$\text{Class Precision Positif} = \frac{61}{88} \times 100 = 69,32\%$$

$$\text{Class Precision Netral} = \frac{53}{68} \times 100 = 77,94\%$$

$$\text{Class Precision Negatif} = \frac{44}{54} \times 100 = 81,48\%$$

Rata-Rata dari *Class Precision* akan dihitung dari total *Class Precision* dan dibagi dengan jumlah kelas yang ada. Rata-rata Precision dihitung sebagai berikut.

$$\text{Precision Avarage} = \frac{\text{Jumlah Precision}}{\text{Jumlah Data Kelas}}$$

$$\text{Precision Avarage} = \frac{69,32\% + 77,94\% + 81,48\%}{3} = 76,25\%$$

Dari hasil di atas maka di ketahui bahwa *precision* dari sistem yang dilakukan adalah 76,25%.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat ditarik kesimpulan bahwa Twitter bisa menjadi sarana untuk mengetahui sentimen masyarakat di Twitter terhadap Danau Toba. Pengambilan data yang mudah juga menjadi alasan kuat untuk mempertimbangkan twitter sebagai sumber data sentimen. Algoritma Naïve Bayes juga bisa memprediksi dengan baik hasil dari analisis sentimen berdasarkan hasil klasifikasi dari sistem analisis yang telah dibuat.

REFERENSI

- [1] Bustami. (2014). PENERAPAN ALGORITMA NAIVE BAYES UNTUK MENGLASIFIKASI DATA NASABAH ASURANSI (Vol. 8, Issue 1).
- [2] Dahri, D., Agus, F., Khairina, D. M., Program,), Komputer, S. I., Komputer, I., Informasi, T., Mulawarman, U., Barong, J., Kampus, T., Samarinda, G. K., & Timur, K. (2016). 5.5 Hal 29 - 36 JIM METODE NAIVE BAYES UNTUK PENENTUAN PENERIMA BEASISWA BIDIKMISI UNIVERSITAS MULAWARMAN. Jurnal Informatika Mulawarman, 11(2), 29.
- [3] Dwi Wijaya, I., Gautama Putrada, A., & Oktaria, D. (n.d.). Penggunaan Metode K- Fold untuk Data Imbalance pada Klasifikasi HWE dan QPQ dalam Kejahatan Tweet Pelecehan Seksual. <https://github.com/amir->
- [4] FAKULTAS ILMU KOMPUTER UNIVERSITAS METHODIST INDONESIA(2018), BUKU PEDOMAN PELAKSANAAN SKRIPSI
- [5] Fridom Mailo, F., Lazuardi, L., Manajemen dan kebijakan Kesehatan Fakultas Kedokteran, D., Masyarakat dan Keperawatan Universitas Gadjah Mada, K., Sistem Informasi Manajemen Kesehatan Fakultas Kedokteran, D., Masyarakat dan Keperawatan, K., & Gadjah Mada, U. (2019). Analisis Sentimen Data Twitter Menggunakan Metode Text

-
- Mining Tentang Masalah Obesitas di Indonesia. In *Jurnal Sistem Informasi Kesehatan Masyarakat Journal of Information Systems for Public Health* (Vol.4,Issue 1).
- [6] Mahardhika, A. A. (2015). NAÏVE BAYES CLASSIFIER DAN COSINE SIMILARITY. 4. <http://www.twitter.com/UPTPuskomUNS>.
- [7] Pandu, A., Nababan, R., Lumenta, A. S. M., Rindengan, Y. D. Y., Pontoh, F. J., Akay, Y. v, Elektro, T., Sam, U., Manado, R., & Manado, J. K. B. (2020). Analisis Sentimen Twitter Pasca Pengumuman Hasil Pilpres 2019 Menggunakan Metode Lexicon Analysis. *Jurnal Teknik Informatika*, 15(1), 33–44. <https://doi.org/10.35793/jti.15.1.2020.29030>
- [8] Perpres No_40 Thn 2016 ttg BADAN OTORITA PENGELOLA DANAU TOBA.pdf
- [9] Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng - Seri Informatika*, 3(1), 50. <https://doi.org/10.35314/isi.v3i1.335>
- [10] Sihombing, R. E., Rachmatin, D., & Dahlan, J. A. (2019). Program Aplikasi Bahasa R Untuk Pengelompokan Objek Menggunakan Metode K-Medoids Clustering. *Program Aplikasi Bahasa R Untuk Pengelompokan Objek Menggunakan Metode K-Medoids Clustering*, 7(1), 58–79.
- [11] Simanjuntak, E. M. (2019). Danau Toba Sebagai Tujuan wisata dari Medan Sumatra Utara. 1–9. <https://doi.org/10.31219/osf.io/6zgbby>
- [12] Suryani, P. S. M., Linawati, L., & Saputra, K. O. (2019). Penggunaan Metode Naive Bayes Classifier pada Analisis Sentimen Facebook Berbahasa Indonesia. *Majalah Ilmiah Teknologi Elektro*, 18(1), <https://doi.org/10.24843/mite.2019.v18i01.p22>
- [13] Thamrin, H. (n.d.). Perancangan Tools Berbasis Python Untuk Memantau 145. Keaktifan Server. Widowati, T. T., & Sadikin, M. (2020). ANALISIS SENTIMEN TWITTER TERHADAP TOKOH PUBLIK DENGAN ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE. *Jurnal SIMETRIS*, 11(2). <https://t.co/Xzf91zHK41>