

Aplikasi Portal Media Sosial Menggunakan Machine Learning Dengan Algoritma Naive Bayes untuk Analisis Sentimen Publik

Raihan Ramadhan Juandri ¹, R. Wisnu Prio Pamungkas ^{1,*}, Ahmad Fathurrozi ¹

* Korespondensi: e-mail: wisnu.prio@dsn.ubharajaya.ac.id

¹ Informatika; Universitas Bhayangkara Jakarta Raya; Jl. Raya Perjuangan, Margamulya, Bekasi; Jawa Barat, Indonesia, Telp.(021)7231948; e-mail: raihan.ramadhan.juandri18@mhs.ubharajaya.ac.id
wisnu.prio@dsn.ubharajaya.ac.id
ahmad.fathurrozi@dsn.ubharajaya.ac.id

Submitted : 29 September 2022
Revised : 21 Oktober 2022
Accepted : 12 November 2022
Published : 30 November 2022

DOI:
<https://doi.org/10.31599/jsracs.v3i2.1760>

Abstract

Many organizations or agencies in Indonesia have several accounts from various social media channels such as YouTube, Facebook, Twitter, and Instagram which are intended as a means of complaining about the products or services they have. As information technology develops, social media continues to grow and requires an organization to create accounts on these channels so that it can continue to relate to the public. Problems like this make it difficult for the organization to process the sentiments conveyed by the public. Therefore, a responsive web-based management system was created that can be operated via a computer or smartphone in categorizing sentiment in the form of data into information using the Naive Bayes method which is fast and high-accuracy in classification by utilizing data sources from social media channels that are integrated in the application.

Keywords: Account, Channels, Naive Bayes, Sentiment, Social Media

Abstrak

Banyak organisasi atau instansi di Indonesia yang memiliki beberapa akun dari berbagai kanal media sosial seperti YouTube, Facebook, Twitter, dan Instagram yang ditujukan sebagai sarana pengaduan mengenai produk atau layanan yang mereka miliki. Semakin berkembangnya teknologi informasi, media sosial terus bertambah dan mengharuskan suatu organisasi untuk membuat akun di kanal-kanal tersebut agar dapat terus terhubung dengan masyarakat. Masalah seperti ini membuat organisasi tersebut menjadi sulit untuk mengolah sentimen yang disampaikan oleh masyarakat. Oleh karena itu dibuatkan sistem pengelolaan berbasis web responsif yang dapat di operasikan melalui komputer maupun ponsel pintar dalam mengkategorikan sentimen berupa data menjadi informasi menggunakan metode Naive Bayes yang berkecepatan dan berketepatan tinggi dalam klasifikasi dengan pemanfaatan sumber data dari kanal media sosial yang terintegrasi dalam aplikasi.

Kata kunci: Akun, Kanal, Media Sosial, Naive Bayes, Sentimen

1. Pendahuluan

Media sosial saat ini memberikan pengaruh besar terhadap lingkungan hidup masyarakat dan mengubah paradigma berkomunikasi yang tidak memiliki batas ruang, waktu, jarak, kapan saja dan dimana saja tanpa harus melalui tatap muka dengan seseorang. Banyak sekali organisasi atau instansi di Indonesia yang memiliki akun dari berbagai kanal media sosial seperti *YouTube*, *Facebook*, *Twitter*, dan *Instagram* dengan tujuan agar dapat terhubung dengan masyarakat terkait produk atau layanan yang mereka miliki. Karena semakin banyak dan berkembangnya kanal media sosial yang beredar membuat organisasi atau instansi tersebut menjadi sulit untuk mengolah sentimen yang disampaikan oleh masyarakat. Oleh karena itu masalah seperti ini dapat diminimalkan dengan dibuatkan sistem pengelolaan yang dapat mengategorikan sentimen-sentimen tersebut yang berupa data menjadi informasi dengan pengelolaan secara responsif dengan kecepatan dan ketepatan yang diintegrasikan menjadi satu dari berbagai kanal media sosial.

Untuk metode pencarian probabilitas suatu kata yang menggunakan pembobotan dengan TF-IDF, banyak sekali penelitian ilmiah yang masih hanya menggunakan persamaan awal (Rizki et al., 2021) yakni:

$$P(X_i|V_j) = \frac{n_k+1}{n+|\text{kosakata}|} \quad (1)$$

Keterangan:

- $P(X_i|V_j)$: Peluang kemunculan X_i pada kategori V_j
 n_k : Jumlah kemunculan frekuensi tiap kata
 n : Jumlah kemunculan kata dari tiap kategori
 $|\text{kosakata}|$: Jumlah semua kata dari kategori yang ada

Persamaan 1 masih dapat dikembangkan jika menggunakan pembobotan dengan TF-IDF untuk mendapatkan hasil yang lebih akurat. Apabila perhitungan probabilitas kata terhadap setiap kelas j menggunakan TF-IDF, maka persamaan 1 disesuaikan menjadi sebagaimana persamaan 2:

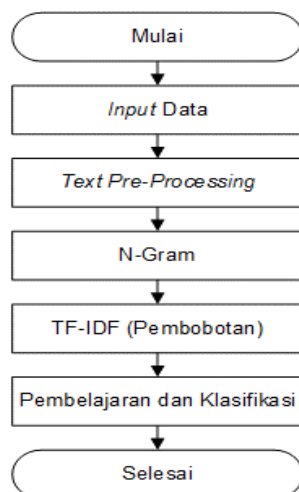
$$P(X_i|V_j) = \frac{\sum_{n=1}^{\infty} (W_{t_n|V_j})+1}{\sum_{n=1}^{\infty} (W_{n|V_j})+\sum_{n=1}^{\infty} (B_n)} \quad (2)$$

Keterangan:

- $P(X_i|V_j)$: Peluang kemunculan X_i pada kategori V_j
 $\sum_{n=1}^{\infty} (W_{t_n|V_j})$: Jumlah bobot (TF-IDF) suatu kata di seluruh dokumen pada kategori V_j
 $\sum_{n=1}^{\infty} (W_{n|V_j})$: Jumlah bobot (TF-IDF) seluruh kata di seluruh dokumen pada kategori V_j
 $\sum_{n=1}^{\infty} (B_n)$: Jumlah seluruh dokumen yang mengandung kosakata atau DF

2. Metode Penelitian

Alur proses penelitian yang dilakukan dengan penerapan menggunakan metode *Naive Bayes* disusun dalam *flowchart* yang dijelaskan pada Gambar 1.



Sumber: Hasil Penelitian (2022)

Gambar 1. *Flowchart* Penelitian

2.1. Metode Penerapan Algoritma Naive Bayes

Sebelum diterapkannya algoritma Naive Bayes, data yang dikumpulkan diproses terlebih dahulu berdasarkan tahap-tahap berikut:

a. *Input Data*

Input data dilakukan dengan menentukan kata kunci yang akan dikumpulkan datanya dari *YouTube* dan *Twitter*. Kata kunci dapat berupa nama pengguna, tagar, maupun kata acak.

b. *Text Pre-processing*

Text preprocessing merupakan suatu tahapan mengubah data yang belum terstruktur menjadi data yang terstruktur (Pramudita et al., 2018). *Text preprocessing* bertujuan untuk memudahkan dalam proses pengolahan data untuk diolah pada tahap selanjutnya. Setelah data didapatkan hasil pada tahap *input data*, dilakukan *text preprocessing* terhadap data yang ada untuk mendapatkan hasil yang lebih akurat saat analisa, meliputi:

- 1) *Cleaning*, pada proses *Cleaning* dilakukan penghapusan *URL (Uniform Resource Locator)*, nama pengguna, tagar, simbol, *punctuation*, karakter angka serta *emoticon* (Rastogi, 2022).
- 2) *Case Folding*, pada tahap *Case Folding* dilakukan perubahan karakter pada data menjadi karakter huruf kecil karena dalam sistem kata yang mengandung huruf besar dengan yang tidak dianggap berbeda (Tineges, 2017), misalnya kata "Coba" dengan C huruf besar, dan "coba" dengan c huruf kecil itu dianggap berbeda.
- 3) *Tokenizing*, pada tahap *Tokenizing*, kata-kata dalam data akan dipisah-pisah dari teks aslinya tanpa memperhatikan adanya duplikasi yang disebut token (Pratama et al., 2021).
- 4) Normalisasi, merupakan proses mengubah kata yang salah eja maupun kata-kata yang disingkat jika ada *dataset*-nya. Misalnya terdapat *dataset* yang memiliki data normalisasi yaitu kata "yang" dengan singkatan "yg". Jika ditemukan kata "yg" pada data maka akan dirubah menjadi "yang".

- 5) *Stopword removal*, pada tahap ini kata yang tidak penting dan yang berulang atau sama akan dihapus (Tineges, 2017), misalnya, “dan”, “di”, “yang”, “lalu”, “karena”, “itu”, dan “oleh”.
- 6) *Stemming*, pada tahap *Stemming* seluruh kata diproses menjadi kata dasar apabila ditemukan kata dasarnya (Guterres et al., 2019), contohnya kata “ngebuat” diubah menjadi kata bakunya yaitu “buat”.

c. *N-Gram*

Proses selanjutnya adalah pemilihan fitur berdasarkan kata menggunakan n-gram. N-gram merupakan salah satu proses yang secara luas digunakan dalam pengolahan teks dan pengolahan bahasa. N-Gram digunakan dalam berbagai macam proses pengolahan bahasa, N-Gram yang digunakan tidak hanya menggunakan *unigram* tapi model *bigram* dan *trigram* juga digunakan. Pada penelitian ini menggunakan N-Gram dengan model *unigram* untuk mengetahui makna dari suatu kalimat maka dilakukan pemotongan kata menggunakan N-Gram.

d. TF-IDF (Pembobotan)

Pada tahap ini dilakukan proses untuk pemberian bobot terhadap kemunculan kata menggunakan metode TF-IDF karena mudah diimplementasikan dan memiliki hasil yang lebih akurat. *Term frequency inverse document frequency* (TF-IDF) sendiri merupakan salah satu metode untuk pembobotan kata (Rofiqi et al., 2019), langkah-langkahnya adalah sebagai berikut:

- 1) Menghitung nilai TF atau kata yang muncul pada data.
- 2) Menghitung nilai IDF menggunakan persamaan berikut:

$$IDF = \log \frac{D}{DF} \quad (3)$$

Keterangan:

IDF : Frekuensi kemunculan kata pada semua dokumen

D : Jumlah dokumen dalam data

DF : Jumlah dokumen yang mengandung kata

- 3) Menghitung nilai akhir TF-IDF menggunakan persamaan berikut:

$$TF-IDF = tf * idf \quad (4)$$

Keterangan:

TF-IDF: Bobot suatu kata

TF : Jumlah kemunculan kata pada dokumen

IDF : Frekuensi kemunculan kata pada semua dokumen

e. Pembelajaran dan Klasifikasi

Dari tahap-tahap yang telah dilakukan sebelumnya, selanjutnya dilakukan proses pembelajaran dan klasifikasi dengan menggunakan algoritma *Naive Bayes*. Algoritma *Naive Bayes* adalah salah satu algoritma klasifikasi berdasarkan teorema Bayesian pada statistika (Suntoro et al., 2018). Algoritma *Naive Bayes* dapat digunakan untuk memprediksi

probabilitas keanggotaan suatu kelas (Suntoro, 2019). Pada metode *Naive Bayes* variabel-variabel yang ada tidak saling terikat antara satu dengan yang lainnya (Akbar et al., 2015). Metode *Naive Bayes* sering digunakan dalam menyelesaikan masalah dalam bidang *machine learning* karena metode ini dikenal memiliki tingkat akurasi yang tinggi dengan perhitungan sederhana (Handayani & Pribadi, 2015). Berikut merupakan persamaan rumus untuk menghitung probabilitas pada masing-masing kelas:

$$P(V_j) = \frac{|dok_j|}{|dok|} \quad (5)$$

Keterangan:

$P(V_j)$: Peluang kemunculan suatu dokumen yang memiliki kategori j

$|dok_j|$: Jumlah dokumen untuk tiap kategori j

$|dok|$: Jumlah dokumen dari semua kategori

Kemudian Untuk melakukan perhitungan probabilitas kata terhadap setiap kelas j dengan menggunakan TF-IDF, maka persamaan yang digunakan adalah persamaan 2. Selanjutnya untuk proses klasifikasi data uji menggunakan persamaan sebagai berikut:

$$V_{MAP} = \arg \max_{V_j \in V} \prod_{i=1}^n P(X_i|V_j)P(V_j) \quad (6)$$

Keterangan:

V_{MAP} : Nilai peluang dokumen terbesar dari semua kategori

V_j : Kategori dokumen $j = 1,2,\dots,n$. Dan dalam penelitian ini

J_1 = kategori negatif, dan

J_2 = kategori positif

$P(X_i|V_j)$: Peluang kemunculan X_i pada kategori V_j

$P(V_j)$: Peluang kemunculan suatu dokumen yang memiliki kategori j

2.2. Metode Pengumpulan Data

Pengumpulan data yang diperlukan untuk kemudian diolah menjadi suatu informasi yang akan di analisa. Pengumpulan data dibagi menjadi 2 (dua) yaitu: (a) Data primer, data yang diperoleh dengan melibatkan kontak langsung atau komunikasi antara peneliti dan informan; (b) Data sekunder, data atau informasi pelengkap data primer yang diambil dari subjek penelitian dalam mendukung kualitas data primer.

3. Hasil dan Pembahasan

Pada algoritma Naive Bayes sebagai analisis sentimen publik berikut langkah dalam penelitian yang dilakukan sesuai dengan urutan algoritma.

3.1. Pengumpulan Data

a. Data Primer (Data Training)

Diawali dengan pengumpulan data sentimen dengan kata kunci acak yang sudah dilabeli kelas atau kategorinya secara manual (Ibrohim et al., 2019).

Tabel 1. Contoh Data Primer (Data Training)

Data Sentimen	Kategori Kelas
Tidak ada solusi lain selain #2019GantiPresiden	Negatif
Ganyang PKI dg segala anteknya... #URL	Negatif
@USER Krn kt percaya, dimanapun kt ditempatkan disitulah kt hrs mengamalkan iman2 kita	Positif
@USER AloClair juga telah dipercaya konsumen di 57 negara, jadi tidak diragukan lagi khasiatnya.	Positif

Sumber: Hasil Penelitian (2022)

b. Data Sekunder (Data Uji)

Pengumpulan data berdasarkan kata kunci yang sudah ditentukan dari media kanal *YouTube* dan *Twitter* menggunakan layanan API setiap kanal.

Tabel 2. Contoh Data Sekunder (Data Uji)

Data Sentimen	Kanal
Bagaimana cara merubah posisi bidang tanah. Yg salah /tidak sesuai lokasi bidang nya..?	<i>Youtube</i>
Kenapa harus ada foto selfie dengan KTP? Ini yang bikin was was mau verifikasi akun.	<i>Youtube</i>
@USER @USER @USER @USER jualan esteh aja pake somasi. pake gula aja kan cukup	<i>Twitter</i>
@USER @USER Sekalian aja surat somasi hahahahaahah	<i>Twitter</i>

Sumber: Hasil Penelitian (2022)

3.2. Perhitungan *Naive Bayes Classifier*

a. Input Data

Contoh data primer dari *Twitter* yang datanya sudah dilindungi kerahasiaannya sesuai kebijakan dari *Twitter* maupun *Youtube* yang mana data primer ditunjukkan pada Tabel 3 yang sudah diketahui kelas atau kategorinya.

Tabel 3. Contoh *Input Data* Primer (Data Training)

Data Sentimen	Kategori
Geblek lo tata...cowo bgt dibela2in balikan...hadeww...ntar ditinggal lg nyalahin tuh cowo...padahal kitenya yg oon.	Negatif
Kmrn termewek2 skr lengket lg duhhh kok labil bgt sih mbak kya abege ajah ato yg kmrn cari sensasi biar top markotoppp ertong gk berkualitas	Negatif
Semangat batman @USER kita VA selalu support kamu karna setiap manusia pasti pernah punya kesalahan ????????, dan buat yang lain tolong jangan di bahas lagi pliiis karna itu udah 2 tahun yang lalu ?????	Positif
USER y udh .salam knl y mbk seneng bisa knl sm mbk. Sbagai manusia kita saling ingetin.y sbagai manusia kita hrs tau diri jgn sombong yg digedein mentang2 dah jd artis haha..y udh mbk kita cabut ja y dri komen d foto ni nnt kita jd korban dy sel	Positif

Sumber: Hasil Penelitian (2022)

Contoh data sekunder dari *Twitter* yang datanya sudah dilindungi kerahasiaannya sesuai kebijakan dari *Twitter* maupun *Youtube* yang mana data sekunder ditunjukkan pada Tabel 4 yang belum diketahui kelas atau kategorinya.

Tabel 4. Contoh *Input Data* Sekunder (Data Uji)

Data Sentimen	Kategori
Senangnya knl sama u, kalo ketemu slalu salam sapa. G pernah sombong	?

Sumber: Hasil Penelitian (2022)

b. Text Pre-processing

Tahap selanjutnya yaitu *Text Pre-processing* pada data di tahap *input data* untuk mendapatkan hasil analisa yang akurat sesuai.

Tabel 5. *Text Pre-Processing Data Primer*

Data Sentimen	Hasil <i>Text Pre-Processing</i>
Geblek lo tata...cowo bgt dibela2in balikan...hadeww...ntar ditinggal lg nyalahin tuh cowo...padahal kitenya yg oon.	geblek, tata, cowok, banget, bela, balik, hadeww, tinggal, salah, cowok, kitenya, beloon
Kmrn termewek2 skr lengket lg duhhh kok labil bgt sih mbak kya abege ajah ato yg kmrn cari sensasi biar top markotopp ertong gk berkualitas	kemarin, mewek, lengket, duhhh, labil, banget, sih, mbak, kayak, abege, kemarin, cari, sensasi, biar, top, markotopp, ertong, kualitas
Semangat batman @USER kita VA selalu support kamu karna setiap manusia pasti pernah punya kesalahan ????????, dan buat yang lain tolong jangan di bahas lagi pliiis karna itu udah 2 tahun yang lalu ????	semangat, batman, va, suppokamu, manusia, salah, tolong, bahas, pliiis
USER y udh .salam knl y mbk seneng bisa knl sm mbk. Sbagai manusia kita saling ingetin.y sbagai manusia kita hrs tau diri jgn sombong yg digedein mentang2 dah jd artis haha..y udh mbk kita cabut ja y dri komen d foto ni nnt kita jd korban dy sel	salam, kenal, mbak, senang, kenal, mbak, manusia, ingat, manusia, sombong, gede, mentang, artis, tertawa, mbak, pergi, komen, foto, korban, sel

Sumber: Hasil Penelitian (2022)

Tabel 6. *Text Pre-Processing Data Sekunder*

Data Sentimen	Kategori
Senangnya knl sama u, kalo ketemu slalu salam sapa. G pernah sombong	senang, kenal, salam, sombong

Sumber: Hasil Penelitian (2022)

c. N-Gram

Setelah melewati tahap *Text Pre-processing*, selanjutnya adalah penentuan N-Gram. Dalam penelitian ini model N-Gram yang digunakan jenis *Unigram*, sehingga hasil yang ditunjukkan pada Tabel 5 dan Tabel 6 bisa digunakan sebagai N-Gram jenis *Unigram*.

d. TF-IDF (Pembobotan)

Proses pembobotan TF-IDF berdasarkan tiap kosakata pada data hasil N-Gram yang berkaitan meliputi:

Tabel 7. TF dan IDF Data Primer dan Data Sekunder

Kata	TF (<i>Term Frequency</i>)					DF	IDF
	D1	D2	D3	D4	D5		
geblek	1	0	0	0	0	1	0,698970004
tata	1	0	0	0	0	1	0,698970004
cowok	2	0	0	0	0	1	0,698970004
banget	1	1	0	0	0	2	0,397940009
bela	1	0	0	0	0	1	0,698970004
balik	1	0	0	0	0	1	0,698970004
hadeww	1	0	0	0	0	1	0,698970004
tinggal	1	0	0	0	0	1	0,698970004
salah	1	0	1	0	0	2	0,397940009
kitenya	1	0	0	0	0	1	0,698970004
beloon	1	0	0	0	0	1	0,698970004
kemarin	0	2	0	0	0	1	0,698970004
mewek	0	1	0	0	0	1	0,698970004
lengket	0	1	0	0	0	1	0,698970004
duhhh	0	1	0	0	0	1	0,698970004
labil	0	1	0	0	0	1	0,698970004
sih	0	1	0	0	0	1	0,698970004
mbak	0	1	0	3	0	2	0,397940009
kayak	0	1	0	0	0	1	0,698970004
abege	0	1	0	0	0	1	0,698970004
cari	0	1	0	0	0	1	0,698970004
sensasi	0	1	0	0	0	1	0,698970004
biar	0	1	0	0	0	1	0,698970004
top	0	1	0	0	0	1	0,698970004
markotoppp	0	1	0	0	0	1	0,698970004
ertong	0	1	0	0	0	1	0,698970004
kualitas	0	1	0	0	0	1	0,698970004
semangat	0	0	1	0	0	1	0,698970004
batman	0	0	1	0	0	1	0,698970004
va	0	0	1	0	0	1	0,698970004
suppokamu	0	0	1	0	0	1	0,698970004
manusia	0	0	1	2	0	2	0,397940009
tolong	0	0	1	0	0	1	0,698970004
bahas	0	0	1	0	0	1	0,698970004
pliiis	0	0	1	0	0	1	0,698970004

Kata	TF (Term Frequency)					DF	IDF
	D1	D2	D3	D4	D5		
salam	0	0	0	1	1	2	0,397940009
kenal	0	0	0	2	1	2	0,397940009
senang	0	0	0	1	1	2	0,397940009
ingat	0	0	0	1	0	1	0,698970004
sombong	0	0	0	1	1	2	0,397940009
gede	0	0	0	1	0	1	0,698970004
mentang	0	0	0	1	0	1	0,698970004
artis	0	0	0	1	0	1	0,698970004
tertawa	0	0	0	1	0	1	0,698970004
pergi	0	0	0	1	0	1	0,698970004
komen	0	0	0	1	0	1	0,698970004
foto	0	0	0	1	0	1	0,698970004
korban	0	0	0	1	0	1	0,698970004
sel	0	0	0	1	0	1	0,698970004

Sumber: Hasil Penelitian (2022)

Tabel 8. TF-IDF Data Primer dan Data Sekunder

Kata	TF-IDF				
	D1	D2	D3	D4	D5
geblek	0,698970004	0	0	0	0
tata	0,698970004	0	0	0	0
cowok	1,397940009	0	0	0	0
banget	0,397940009	0,397940009	0	0	0
bela	0,698970004	0	0	0	0
balik	0,698970004	0	0	0	0
hadeww	0,698970004	0	0	0	0
tinggal	0,698970004	0	0	0	0
salah	0,397940009	0	0,397940009	0	0
kitenya	0,698970004	0	0	0	0
beloon	0,698970004	0	0	0	0
kemarin	0	1,397940009	0	0	0
mewek	0	0,698970004	0	0	0
lengket	0	0,698970004	0	0	0
duhhh	0	0,698970004	0	0	0
labil	0	0,698970004	0	0	0
sih	0	0,698970004	0	0	0
mbak	0	0,397940009	0	1,193820026	0
kayak	0	0,698970004	0	0	0
abege	0	0,698970004	0	0	0

Kata	TF-IDF				
	D1	D2	D3	D4	D5
cari	0	0,698970004	0	0	0
sensasi	0	0,698970004	0	0	0
biar	0	0,698970004	0	0	0
top	0	0,698970004	0	0	0
markotoppp	0	0,698970004	0	0	0
ertong	0	0,698970004	0	0	0
kualitas	0	0,698970004	0	0	0
semangat	0	0	0,698970004	0	0
batman	0	0	0,698970004	0	0
va	0	0	0,698970004	0	0
suppokamu	0	0	0,698970004	0	0
manusia	0	0	0,397940009	0,795880017	0
tolong	0	0	0,698970004	0	0
bahas	0	0	0,698970004	0	0
pliiis	0	0	0,698970004	0	0
salam	0	0	0	0,397940009	0,397940009
kenal	0	0	0	0,795880017	0,397940009
senang	0	0	0	0,397940009	0,397940009
ingat	0	0	0	0,698970004	0
sombong	0	0	0	0,397940009	0,397940009
gede	0	0	0	0,698970004	0
mentang	0	0	0	0,698970004	0
artis	0	0	0	0,698970004	0
tertawa	0	0	0	0,698970004	0
pergi	0	0	0	0,698970004	0
komen	0	0	0	0,698970004	0
foto	0	0	0	0,698970004	0
korban	0	0	0	0,698970004	0
sel	0	0	0	0,698970004	0

Sumber: Hasil Penelitian (2022)

e. Pembelajaran dan Klasifikasi

Untuk mengetahui hasil klasifikasi pada data sekunder atau data uji, maka tahap-tahap yang dilakukan adalah:

- 1) Mencari probabilitas tiap kelas atau kategori dari data primer atau data yang sudah diketahui kelasnya menggunakan persamaan 5.

$$P(\text{Positif}) = \frac{2}{4} = 0.5$$

$$P(\text{Negatif}) = \frac{2}{4} = 0.5$$

- 2) Setelah mencari probabilitas tiap kelas, selanjutnya mencari probabilitas setiap kata pada data uji menggunakan persamaan 2 berdasarkan kelas atau kategori yang dicari.

Tabel 9. Probabilitas Kata Kelas Positif

Probabilitas	Persamaan	Hasil
$P(\text{Senang} \text{Positif})$	$= \frac{\sum_{n=1}^{\infty} (W_{\text{Senang}_n} \text{Positif})+1}{\sum_{n=1}^{\infty} (W_n \text{Positif}) + \sum_{n=1}^{\infty} (B_n)}$	$= \frac{0,79588+1}{18,24953+ 31,84129} = 0,035852$
$P(\text{Kenal} \text{Positif})$	$= \frac{\sum_{n=1}^{\infty} (W_{\text{Kenal}_n} \text{Positif})+1}{\sum_{n=1}^{\infty} (W_n \text{Positif}) + \sum_{n=1}^{\infty} (B_n)}$	$= \frac{1,19382+1}{18,24953+ 31,84129} = 0,043797$
$P(\text{Salam} \text{Positif})$	$= \frac{\sum_{n=1}^{\infty} (W_{\text{Salam}_n} \text{Positif})+1}{\sum_{n=1}^{\infty} (W_n \text{Positif}) + \sum_{n=1}^{\infty} (B_n)}$	$= \frac{0,79588+1}{18,24953+ 31,84129} = 0,035852$
$P(\text{Sombong} \text{Positif})$	$= \frac{\sum_{n=1}^{\infty} (W_{\text{Sombong}_n} \text{Positif})+1}{\sum_{n=1}^{\infty} (W_n \text{Positif}) + \sum_{n=1}^{\infty} (B_n)}$	$= \frac{0,79588+1}{18,24953+ 31,84129} = 0,035852$

Sumber: Hasil Penelitian (2022)

Tabel 10. Probabilitas Kata Kelas Negatif

Probabilitas	Persamaan	Hasil
$P(\text{Senang} \text{Negatif})$	$= \frac{\sum_{n=1}^{\infty} (W_{\text{Senang}_n} \text{Negatif})+1}{\sum_{n=1}^{\infty} (W_n \text{Negatif}) + \sum_{n=1}^{\infty} (B_n)}$	$= \frac{0,39794+1}{21,35674+ 31,84129} = 0,026278$
$P(\text{Kenal} \text{Negatif})$	$= \frac{\sum_{n=1}^{\infty} (W_{\text{Kenal}_n} \text{Negatif})+1}{\sum_{n=1}^{\infty} (W_n \text{Negatif}) + \sum_{n=1}^{\infty} (B_n)}$	$= \frac{0,39794+1}{21,35674+ 31,84129} = 0,026278$
$P(\text{Salam} \text{Negatif})$	$= \frac{\sum_{n=1}^{\infty} (W_{\text{Salam}_n} \text{Negatif})+1}{\sum_{n=1}^{\infty} (W_n \text{Negatif}) + \sum_{n=1}^{\infty} (B_n)}$	$= \frac{0,39794+1}{21,35674+ 31,84129} = 0,026278$
$P(\text{Sombong} \text{Negatif})$	$= \frac{\sum_{n=1}^{\infty} (W_{\text{Sombong}_n} \text{Negatif})+1}{\sum_{n=1}^{\infty} (W_n \text{Negatif}) + \sum_{n=1}^{\infty} (B_n)}$	$= \frac{0,39794+1}{21,35674+ 31,84129} = 0,026278$

Sumber: Hasil Penelitian (2022)

- 3) Tahap selanjutnya yaitu mengalikan semua nilai probabilitas berdasarkan kelas atau kategori dengan tujuan untuk mengetahui kelas atau kategori data uji menggunakan persamaan 6.

$$\begin{aligned}
 P(\text{Data Uji}|\text{Positif}) &= P(\text{Senang, Kenal, Salam, Sombong}|\text{Positif})P(\text{Positif}) \\
 &= 0,035852 \times 0,043797 \times 0,035852 \times 0,035852 \times 0,5 \\
 &= 1,01E-06 \text{ atau } 0,00000101
 \end{aligned}$$

$$\begin{aligned}
 P(\text{Data Uji}|\text{Negatif}) &= P(\text{Senang, Kenal, Salam, Sombong}|\text{Negatif})P(\text{Negatif}) \\
 &= 0,026278 \times 0,026278 \times 0,026278 \times 0,026278 \times 0,5 \\
 &= 2,38E-07 \text{ atau } 0,000000238
 \end{aligned}$$

$$\begin{aligned}
 V_{\text{MAP}} &= \{ \text{Positif} = 0,00000101, \text{Negatif} = 0,000000238 \} \\
 &= \text{Positif}
 \end{aligned}$$

Dari hasil V_{MAP} diketahui bahwa hasil klasifikasi data uji tertinggi adalah kelas Positif, sehingga dapat dipastikan kelas atau kategori data uji adalah kelas Positif.

4. Kesimpulan

Berdasarkan hasil implementasi algoritma *Naive Bayes* dalam mengklasifikasikan sentimen dari kanal media sosial dapat disimpulkan bahwa sistem aplikasi Portal Media Sosial

mampu membantu instansi atau organisasi dalam mengelola komentar-komentar yang dikirimkan oleh masyarakat atau konsumen dalam berbagai perangkat, dengan menggunakan algoritma *Naive Bayes* serta pelayanan menjadi lebih efektif dan efisien dengan adanya sistem aplikasi yang mengklasifikasikan sentimen-sentimen secara otomatis.

Daftar Pustaka

- Akbar, A. S., Sedyono, E., & Nurhayati, O. D. (2015). Analisis Sentimen Berbasis Ontologi di Level Kalimat untuk Mengukur Persepsi Produk. *JSINBIS (Jurnal Sistem Informasi Bisnis)*, 5(2), 84–97. <https://doi.org/10.21456/vol5iss2pp84-97>
- Guterres, A., Gunawan, & Santoso, J. (2019). Stemming Bahasa Tetun Menggunakan Pendekatan Rule Based. *Teknika*, 8(2), 142–147.
- Handayani, F., & Pribadi, F. S. (2015). Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110. *Jurnal Teknik Elektro*, 7(1), 19–24.
- Ibrohim, Okky, M., Budi, & Indra. (2019). Multi-label Hate Speech and Abusive Language Detection in Indonesian Twitter. *Proceedings of the Third Workshop on Abusive Language Online*, 46–57.
- Pramudita, Y. D., Putro, S. S., & Makhmud, N. (2018). Klasifikasi Berita Olahraga Menggunakan Metode Naive Bayes dengan Enhanced Confix Stripping Stemmer. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 5(3), 269–276.
- Pratama, W., Ilyas, R., & Kasyidi, F. (2021). Peringkasan Otomatis Makalah Menggunakan Maximum Marginal. *Informatics And Digital Expert (INDEX)*, 3(1), 32–37. <https://doi.org/10.36423/index.v3i1.677>
- Rastogi, K. (2022). *Text Cleaning Methods in NLP*. <https://www.analyticsvidhya.com/blog/2022/01/text-cleaning-methods-in-nlp/>
- Rizki, M., Arhami, M., & Huzeni. (2021). Perbaikan Algoritma Naive Bayes Classifier Menggunakan Teknik Laplacian Correction. *Jurnal Teknologi*, 21(1), 39–45.
- Rofiqi, M. A., Fauzan, A. C., Agustin, A. P., Saputra, A. A., & Fahma, H. D. (2019). Implementasi Term-Frequency Inverse Document Frequency (TF-IDF) Untuk Mencari Relevansi Dokumen Berdasarkan Query. *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, 1(2), 58–64.
- Suntoro, J. (2019). *Data Mining: Algoritma dan Implementasi dengan Pemrograman PHP*. Elex Media Komputindo.
- Suntoro, J., Christanto, F. W., & Indriyawati, H. (2018). Software Defect Prediction Using AWEIG+ADACOST Bayesian Algorithm for Handling High Dimensional Data and Class Imbalance Problem. *International Journal of Information Techonology and Business*, 1(1), 36–41.
- Tineges, R. (2017). *Tahapan Text Preprocessing dalam Teknik Pengolahan Data*. <https://www.dqlab.id/tahapan-text-preprocessing-dalam-teknik-pengolahan-data>