

ANALISIS SENTIMEN PADA KOMENTAR INSTAGRAM MENGGUNAKAN GAUSSIAN NAÏVE BAYES**M Fikri Rizki Pratama¹, M Ezar Al Rivani²**¹Fakultas Ilmu Komputer dan Rekayasa, Informatika, Universitas Multi Data Palembang, Palembang, IndonesiaEmail: fr676824@mhs.mdp.ac.id, meedzhar@mdp.ac.id, fr676824@mhs.mdp.ac.id**Abstract (English)**

In general, sentiment analysis is a way to identify public opinion expressed in a text. Opinions have the form of emotions, evaluations, beliefs, satisfaction and judgments. Social Media is a medium on the internet that allows users to present themselves and interact, collaborate, share, communicate with other users, and form virtual social bonds. Instagram is a social media platform that allows users to share photos and videos, as well as interact with other users through comments. In this study, the Naive Bayes classification method was used using a dataset of 400. The dataset was divided into training data and testing data with a ratio of 80%: 20%. The results of this research showed that the negative class produced a precision value of 82%, Recall 78%, F1-Score 79%. while the positive class produces a precision value of 79%, Recall 82%, F1-Score 80%.

Abstrak (Indonesia)

Secara umum analisis sentimen adalah cara untuk mengenali opini publik yang dituangkan dalam sebuah teks. Opini memiliki bentuk emosi, evaluasi, kepercayaan, kepuasan dan penilaian. Media Sosial merupakan medium di internet yang memungkinkan pengguna mempresentasikan dirinya maupun berinteraksi, bekerjasama, saling berbagi, berkomunikasi dengan pengguna lainnya, dan membentuk ikatan sosial secara virtual. Instagram adalah sebuah platform media sosial yang memungkinkan pengguna untuk berbagi foto dan video, serta berinteraksi dengan pengguna lainnya melalui komentar. Pada penelitian ini menggunakan metode klasifikasi naive bayes dengan menggunakan dataset yang berjumlah 400. Dataset dibagi menjadi data training dan data testing dengan perbandingan rasio 80%: 20%. Hasil dari penelitian ini mendapatkan kelas negative menghasilkan nilai precision 82%, Recall 78%, F1-Score 79%. sedangkan untuk kelas positive menghasilkan nilai precision 79%, Recall 82%, F1-Score 80%.

Article History

Submitted: 6 Desember 2024
Accepted: 9 Desember 2024
Published: 16 Desember 2024

Key Words

Sentiment Analysis; Naive Bayes Classification; Instagram; Social media

Sejarah Artikel

Submitted: 6 Desember 2024
Accepted: 9 Desember 2024
Published: 16 Desember 2024

Kata Kunci

Analisis Sentimen; Klasifikasi Naive Bayes; Instagram; Media Sosial.

1. PENDAHULUAN

Analisis sentimen merupakan proses memahami, mengekstrak dan mengolah data tekstual untuk mendapatkan informasi yang terkandung dalam suatu kalimat. Secara umum analisis sentimen adalah cara untuk mengenali opini publik yang dituangkan dalam sebuah teks. Opini memiliki bentuk emosi, evaluasi, kepercayaan, kepuasan dan penilaian. Analisis sentiment digunakan untuk melihat pendapat publik terhadap suatu masalah. (M Indra Halim Arsyah Dwi Akbari et al., 2012)

Menurut Nasrullah (2016:13), "Media Sosial merupakan medium di internet yang memungkinkan pengguna mempresentasikan dirinya maupun berinteraksi, bekerjasama, saling berbagi, berkomunikasi dengan pengguna lainnya, dan membentuk ikatan sosial secara virtual."

Instagram adalah sebuah platform media sosial yang memungkinkan pengguna untuk berbagi foto dan video, serta berinteraksi dengan pengguna lainnya melalui komentar, pesan langsung, dan fitur-fitur lainnya. Instagram memiliki jumlah pengguna aktif yang sangat besar, termasuk di Indonesia, dan telah menjadi bagian dari kehidupan sehari-hari banyak orang, terutama generasi milenial dan generasi Z. (Wanda Athira Luqyana et al., 2018) Komentar telah menjadi salah satu bentuk interaksi sosial media yang populer di kalangan pengguna internet saat ini. Namun, karena jumlah komentar yang sangat banyak, sulit bagi

manusia untuk secara efisien memonitor dan mengidentifikasi emosi dalam setiap komentar. Oleh karena itu, deteksi emosi pada teks komentar Instagram dapat menjadi masalah yang menarik untuk diselesaikan menggunakan teknik-teknik pemrosesan bahasa alami. (Wanda Athira Luqyana et al., 2018)

♦ Algoritma Klasifikasi Naive Bayes adalah metode klasifikasi yang berdasarkan teorema Bayes. Metode ini menggunakan metode probabilitas dan statistik yang ditemukan oleh ilmuwan Inggris bernama Thomas Bayes, yaitu sebuah metode untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya, sehingga metode ini dikenal sebagai Teorema Bayes. Klasifikasi Naive Bayes memiliki ciri khas asumsi yang sangat kuat akan independensi dari masing-masing kondisi atau kejadian. (M. Syukri Mustafa et al., 2017)

Pada penelitian tahun 2020 yang dilakukan oleh Reski Mai Candra dan Anindya Nanda Rozana dengan judul Klasifikasi Komentar Bullying Instagram Menggunakan Metode K-Nearest Neighbor dengan menggunakan dataset dari Instagram yang berjumlah 1000 data, penelitian ini menggunakan metode K-NN menghasilkan nilai accuracy rata-rata sebesar 58,83%.

Pada penelitian tahun 2018 yang dilakukan oleh Wanda Athira Luqyana et al., dengan judul Analisis Sentimen CyberBullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. dataset digunakan berasal dari Instagram, penelitian ini menggunakan Metode Klasifikasi Support Vector Machine dan Term Frequency-Inverse Document Frequency (TF-IDF) dan menghasilkan nilai akurasi tertinggi sebesar 90% pada komposisi data latih 50% dan data uji 50%.

Pada penelitian yang dilakukan oleh Sepyan Purnama Kristanto et al., pada tahun 2019 dengan judul Naïve Bayes Classifier on Twitter Sentiment Analysis BPJS of Health data yang digunakan berasal dari twitter, penelitian ini menggunakan Naïve Bayes dan menghasilkan nilai akurasi sebesar 70% untuk topik pembayaran negatif dan 72% untuk topik pembayaran positif dan 65% untuk pengguna yang menggunakan bpjs sebagai layanan kesehatan.

Pada penelitian tahun 2020 yang dilakukan oleh Angelina Puput Giovani et al., dengan judul Analisis Sentimen Aplikasi Ruang Guru di Twitter Menggunakan Algoritma Klasifikasi dataset yang digunakan berasal dari Twitter. Penelitian ini menggunakan Algoritma Naïve Bayes, Support Vector Machine, K-Nearest Neighbour dan feature Selection dengan menggunakan Algoritma Particle Swarm Optimization dan mendapatkan hasil terbaik menggunakan Algoritma PSO berbasis SVM dengan nilai akurasi 78,55%.

Pada yang dilakukan oleh Ni Luh Putu Merawati et al., pada tahun 2021 dengan judul penelitian Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation. Dataset yang digunakan berasal dari postingan user twitter. Penelitian ini menggunakan metode Naive Bayes dan penggunaan metode Latent Dirichlet Allocation dengan menghasilkan nilai akurasi 92%, nilai presisi 100%, nilai recall 84%, nilai spesifitas 100%. Nilai koherensi kelas positif mendapatkan nilai sebesar 0,613 sedangkan untuk Nilai koherensi kelas negatif sebesar 0,528.

Pada penelitian tahun 2022 yang dilakukan oleh Slamet Harry Ramadhani dan Muhammad Iwan Wahyudin dengan judul Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter. Penelitian ini menggunakan Metode Naive Bayes dan K-NN. Dataset yang digunakan berasal dari postingan twitter. Penelitian ini menghasilkan nilai akurasi 90.71% untuk Naive Bayes sedangkan untuk nilai akurasi K-NN sebesar 74.78%.

Berdasarkan pada penelitian yang telah dilakukan. Penelitian ini akan membahas tentang bagaimana implementasi algoritma klasifikasi Naive Bayes pada teks berbahasa Indonesia dapat membantu dalam analisis sentimen pada teks. Selain itu, penelitian ini juga akan mengidentifikasi tantangan yang dihadapi dalam analisis sentimen pada teks berbahasa Indonesia dan bagaimana cara mengatasi tantangan tersebut.

Diharapkan hasil dari penelitian ini dapat membantu pengguna media sosial dan bisnis online dalam memahami sentimen pengguna dan mengambil tindakan yang tepat untuk meningkatkan interaksi dan kepuasan pengguna. Selain itu, penelitian ini juga dapat menjadi kontribusi dalam pengembangan bidang analisis sentimen pada teks berbahasa Indonesia.

2. METODE PENELITIAN

2.1 Analisis Sentimen

Analisis sentimen merupakan cabang dari ilmu Natural Language Processing dengan melakukan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat untuk menjadi sebuah informasi yang bermanfaat. Informasi yang didapat dari analisis sentimen bisa menjadi acuan penilaian terhadap suatu peristiwa, topik, organisasi, maupun perorangan. (Wanda Athira Luqyana et al., 2018)

2.2 Algoritma Naïve Bayes Classifier

Naive Bayes Classifier adalah sebuah metode klasifikasi yang berdasar pada teorema Bayes. Metode pengklasifikasian ini menggunakan metode probabilitas dan statistik yang pertama kali dikemukakan oleh ilmuwan Inggris bernama Thomas Bayes, yaitu suatu metode untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya, sehingga metode ini dikenal sebagai Teorema Bayes. Ciri utama dari Naive Bayes Classifier ini adalah asumsi yang sangat kuat akan independensi dari masing-masing kondisi atau kejadian. (M. Syukri Mustafa et al., 2017)

$$P(H|X) = \frac{P(X|H) P(H)}{P(X)} \dots\dots (1)$$

Keterangan :

X : data dengan class yang belum diketahui

H : hipotesis data x merupakan suatu kelas spesifik

P(H|X) : probabilitas hipotesis H berdasar kondisi X

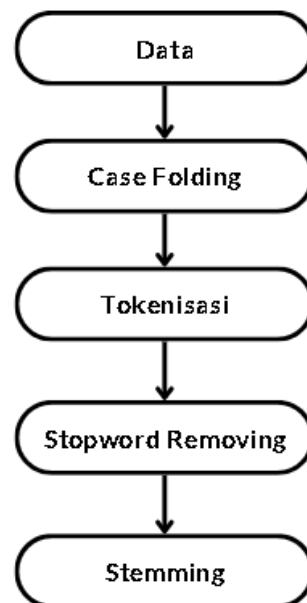
P(H) : probabilitas hipotesis H

P(X|H): probabilitas hipotesis X berdasar kondisi H

P(X) : probabilitas dari X

2.3 Pre-Processing

Pre-processing merupakan langkah untuk bahasa yang tidak baku menjadi baku dan membersihkan data yang ada sehingga data menjadi lebih rapi dan terstruktur untuk dipakai.



Gambar 1. Proses Pre-Processing

2.3.1 Case Folding

Case Folding adalah proses mengubah semua huruf dalam teks menjadi huruf kecil atau huruf besar. Hal ini dilakukan agar tidak terjadi duplikasi token karena adanya perbedaan huruf besar dan huruf kecil. Contohnya, kata "Hello" dan "hello" akan dianggap sebagai token yang berbeda.

2.3.2 Tokenisasi

Tokenisasi adalah proses memecah teks menjadi bagian-bagian yang lebih kecil, yang disebut sebagai token. Token dapat berupa kata, frasa, atau tanda baca. Tokenisasi biasanya dilakukan dengan memisahkan setiap kata dalam teks menggunakan spasi sebagai pemisah. Contohnya, kalimat "Saya suka makan nasi goreng" akan dibagi menjadi token "Saya", "suka", "makan", "nasi", dan "goreng".

2.3.3 Stopword Removing

Stopword Removing adalah proses menghilangkan kata-kata yang umum dan tidak memiliki makna khusus dalam teks. Kata-kata ini sering disebut sebagai stopword, seperti "dan", "atau", "yang", dan "di". Stopword sering dihilangkan karena tidak memberikan kontribusi yang signifikan dalam analisis teks dan hanya membuang-buang sumber daya komputasi. Contohnya, jika kalimat "Saya dan teman-teman saya makan di restoran" diproses dengan stopword removal, maka kata "dan" akan dihapus dan kalimat tersebut menjadi "Saya teman-teman makan restoran".

2.3.4 Stemming

Stemming adalah proses mengubah kata-kata dalam teks menjadi kata dasarnya dengan menghilangkan imbuhan atau awalan pada kata. Tujuan dari stemming adalah untuk mengurangi variasi token dalam teks dan menghasilkan representasi yang lebih sederhana dan konsisten. Contohnya kata "berlari", "berlarian", dan "berlariilah" akan diubah menjadi kata dasarnya "lari".

2.4 Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF adalah suatu metode yang bisa digunakan untuk pembobotan kata. Term Weighting atau pembobotan kata bertujuan untuk memberikan bobot nilai pada setiap kata.

Perhitungan bobot ini memerlukan dua hal yaitu Term Frequency (TF) dan Inverse Document Frequency (IDF) (Amelia Rahman et al., 2017)

Term Frequency merupakan banyaknya jumlah kata atau term tertentu yang ada dalam suatu dokumen. Sementara Inverse Document Frequency adalah frekuensi kemunculan kata atau term pada seluruh dokumen. Sebelum menghitung bobot TF-IDF terlebih dahulu menghitung bobot IDF (Mega Kurnia Maulidina & Enny Itje Sela, 2020)

2.5 Gaussian Naive Bayes

Gaussian Naive Bayes adalah suatu model yang memungkinkan untuk mengklasifikasi data dengan distribusi Gaussian atau distribusi normal. Distribusi Gaussian mengasumsikan bahwa setiap feature data memiliki pengaruh yang sendiri dalam memprediksi data. Rumus untuk menghitung probabilitas menggunakan Gaussian Naive Bayes sebagai berikut (Nur Ghaniaviyanto Ramadhan & Faisal Dharma Adhinata, 2022) rumus Gaussian Naive Bayes sebagai berikut :

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}} \dots\dots (2)$$

Keterangan

- P : Nilai peluang
- Xi : Nilai atribut ke- i
- y : Kelompok y
- μy : Rata-rata kelompok y
- σy : Variasi kelompok ke y

2.6 Sastrawi

Sastrawi adalah sebuah library atau perpustakaan open source yang digunakan untuk melakukan stemming pada bahasa Indonesia. Stemming merupakan suatu teknik dalam pengolahan teks yang bertujuan untuk mengubah kata-kata dalam bentuk infleksional atau varian kata menjadi kata dasar atau kata dasarnya. mengambil kata dasar dari setiap kata-kata yang menjadi data masukan. Instalasi Sastrawi menggunakan PHP Composer sehingga library Sastrawi dapat terunduh dan digunakan. Sastrawi juga dapat digunakan untuk proses stopword removal, yaitu menghilangkan kata stopword. Sastrawi sudah dilengkapi dengan daftar kata dalam bahasa Indonesia dan daftar kata stopword dalam bahasa Indonesia. (Antonius Rachmat C & Yuan Lukito, 2016)

3. HASIL DAN PEMBAHASAN

3.1 Hasil

3.1.1 Case Folding

Tahapan pertama dalam proses *pre-processing* adalah *case folding*. Dalam proses pertama dataset akan dilakukan transformasi huruf dari huruf besar menjadi huruf kecil. Hasil dari proses case folding bisa dilihat di tabel 1.

Tabel 1 Case Folding

Sebelum	<USERNAME> TOLOL!! Gak ada hubungan nya keguguran dgn pake hijab syar'i yg lo bilang bayi nya kepanasan didalam gak ada hubungan nya woyyyy!! Otak sama jempol lo gak sinkron sih ya jadinya asal nulis komentar!
Sesudah	<username> tolol!! gak ada hubungan nya keguguran dgn pake hijab syar'i yg lo bilang bayi nya kepanasan didalam gak ada hubungan nya woyyyy!! otak sama jempol lo gak sinkron sih ya jadinya asal nulis komentar!

3.1.2 Tokenisasi

Setelah dilakukan case folding maka dilanjutkan dengan tahapan berikutnya yaitu tahapan tokenisasi. Tokenisasi dilakukan untuk menghapus tanda baca, angka dan *emoticon* yang ada pada teks. Serta memisahkan antar kata pada kalimat. Hasil dari proses Tokenisasi bisa dilihat di tabel 2.

Tabel 2 Tokenisasi

Sebelum	<USERNAME> TOLOL!! Gak ada hubungan nya keguguran dgn pake hijab syar'i yg lo bilang bayi nya kepanasan didalem gak ada hubungan nya woyyyyy!! Otak sama jempol lo gak sinkron sih ya jadinya asal nulis komentar!
Sesudah	'username', 'tolol', 'hubungan', 'keguguran', 'pake', 'hijab', 'syar', 'bilang', 'bayi', 'kepanasan', 'didalem', 'hubungan', 'woyyyyy', 'otak', 'sama', 'jempol', 'sinkron', 'jadinya', 'asal', 'nulis', 'komentar'

3.1.3 Stopword Removing

Pada Tahapan tiga akan dilakukan proses *Stopword Removing*. Pada tahapan ini dilakukan untuk menghapus kata kata tidak memberikan kontribusi yang signifikan dalam analisis teks dan hanya membuang-buang sumber daya komputasi.

Pada tahapan *Stopword Removing* akan menggunakan NLTK yang sudah disediakan oleh bahasa pemrograman Python. Hasil dari proses *Stopword Removing* bisa dilihat di tabel 3.

Tabel 3 Stopword Removing

Sebelum	<USERNAME> TOLOL!! Gak ada hubungan nya keguguran dgn pake hijab syar'i yg lo bilang bayi nya kepanasan didalem gak ada hubungan nya woyyyyy!! Otak sama jempol lo gak sinkron sih ya jadinya asal nulis komentar!
Sesudah	'username', 'tolol', 'hubungan', 'keguguran', 'pake', 'hijab', 'syar', 'bilang', 'bayi', 'kepanasan', 'didalem', 'hubungan', 'woyyyyy', 'otak', 'jempol', 'sinkron', 'nulis', 'komentar'

3.1.4 Stemming

Proses *Stemming* merupakan tahapan terakhir pada *pre-processing*. Pada tahapan ini akan dilakukan proses perubahan kata-kata dalam teks yang memiliki imbuhan menjadi kata dasar

Pada tahapan *Stemming* akan menggunakan modul Sastrawi yang sudah disediakan oleh bahasa pemrograman Python. Hasil dari proses *Stemming* bisa dilihat di tabel 4.

Tabel 4 Stemming

Sebelum	<USERNAME> TOLOL!! Gak ada hubungan nya keguguran dgn pake hijab syar'i yg lo bilang bayi nya kepanasan didalem gak ada hubungan nya woyyyyy!! Otak sama jempol lo gak sinkron sih ya jadinya asal nulis komentar!
Sesudah	'username', 'tolol', 'hubung', 'gugur', 'pake', 'hijab', 'syar', 'bilang', 'bayi', 'panas', 'dalem', 'hubung', 'woyyyyy', 'otak', 'jempol', 'sinkron', 'nulis', 'komentar'

3.1.5 Term Frequency-Inverse Document Frequency (TF-IDF)

Setelah melakukan *pre-processing* maka tahapan berikutnya akan dilakukan pembobotan pada setiap kata. Tahapan ini dilakukan untuk mengetahui nilai dari setiap kata yang sudah melalui tahapan *pre-processing*.

Pada penelitian ini, proses TF-IDF dilakukan dengan menggunakan modul scikit-learn yang disediakan oleh bahasa pemrograman python. Hasil dari pembobotan TF-IDF dari 400 data komentar instagram bisa dilihat pada tabel 5.

Tabel 5 Hasil TF-IDF

	aamiin	aammiinnn	abege	...	yess	yesss	yesssssss
0	0.0	0.0	0.0	...	0.0	0.0	0.0
1	0.0	0.0	0.0	...	0.0	0.0	0.0
2	0.0	0.0	0.0	...	0.0	0.0	0.0
3	0.0	0.0	0.0	...	0.0	0.0	0.0
4	0.0	0.0	0.0	...	0.0	0.0	0.0
...
395	0.0	0.0	0.0	...	0.0	0.0	0.0
396	0.0	0.0	0.0	...	0.0	0.0	0.0
397	0.0	0.0	0.0	...	0.0	0.0	0.0
398	0.0	0.0	0.0	...	0.0	0.0	0.0
399	0.0	0.0	0.0	...	0.0	0.0	0.0

```

(0, 883)      0.256978987740306
(0, 977)      0.21053169898980198
(0, 42)       0.1617985484624172
(0, 1357)     0.513957975480612
(0, 1065)     0.513957975480612
(0, 1548)     0.16650096190136196
(0, 361)      0.21823991605709575
(0, 385)      0.2040144943208979
(0, 299)      0.256978987740306
(0, 1249)     0.23983667543521978
(0, 972)      0.16906405840405248
(0, 1571)     0.256978987740306
(1, 1019)     0.46920871109878143
(1, 935)      0.3844022734974292
(1, 401)      0.46920871109878143
(1, 243)      0.46920871109878143
(1, 104)      0.4379091782745235

```

Gambar 2 Matriks TF-IDF

Matriks TF-IDF merupakan matriks sparse atau banyak mengandung nilai 0 didalamnya, jika diubah kedalam bentuk barisan maka diperoleh hasil untuk semua nilai TF-IDF > 0 seperti yang ditunjukkan oleh Gambar 2. Kemudian, sebagai contoh, Langkah-langkah melakukan proses Tf-Idf :

1. Term Frequency (TF)

Proses TF dilakukan untuk mengukur seberapa sering kata muncul dalam satu dokumen tertentu. Semakin sering kata muncul, semakin tinggi nilai TF-nya. komentar : “senang liat nikmat brsama cucu cinta sehat memo pepo allah lindung memo pepo kiss” Jumlah total kata dalam komentar ada 14 dan menghitung kata kiss yang 1 kali , maka :

$$TF(kiss) = \frac{1}{14} = 0,07142$$

2. Inverse Document Frequency (IDF)

Proses IDF dilakukan untuk mengukur seberapa umum atau jarang suatu kata muncul di seluruh komentar dalam dataset. Semakin jarang kata muncul di dokumen lain, semakin tinggi nilai IDFnya, menunjukkan bahwa kata tersebut memiliki nilai informasi yang tinggi.

$$IDF(kiss) = \log\left(\frac{320}{1}\right) = 2,50514$$

3. TF-IDF

Setelah menghitung nilai TF dan IDF selanjutnya menghitung nilai TF-IDF

$$Tf - Idf(kiss) = 0,07142 \times 2,50514 = 0,17891$$

4. Normalisasi

Normalisasi dilakukan untuk mengatasi perbedaan panjang komentar sehingga model yang digunakan bekerja lebih efektif

$$L2 = \sqrt{\frac{(0,25)^2 \times (0,21)^2 \times (0,16)^2 \times (0,51)^2 \times (0,51)^2}{x (0,16)^2 \times (0,21)^2 \times (0,20)^2 \times (0,25)^2 \times (0,23)^2}} = 0,7814$$

$$Tf - Idf \text{ normalisasi } (kiss) = \frac{0,17891}{0,7814} = 0.256978987740306$$

Setelah melakukan proses tf-idf mendapatkan nilai pada setiap kata dalam dokumen, jika nilai tf-idf adalah 0.0 maka kata tersebut tidak memberikan informasi yang berguna untuk membedakan komentar satu sama komentar yang lain dalam dataset, jika nilai tf-idf lebih besar dari 0.0 menunjukan kata tersebut ada dalam komentar tertentu dan tidak ada di setiap komentar dalam dataset, semakin tinggi nilai tf-idf maka semakin penting kata tersebut untuk membedakan komentar satu sama komentar yang lain dalam dataset .

3.1.6 Model Gaussian Naive Bayes

Pada tahapan klasifikasi, data yang sudah melalui tahapan *pre-processing* dan pembobotan oleh TF-IDF dibagi menjadi data training dan data testing dengan perbandingan rasio 80% data training: 20% data testing. Pada tahap training model yang dihasilkan oleh Gaussian Naive Bayes ditunjukkan oleh Gambar 3.

```
[15] print(model.theta_)
      print(model.theta_.shape)

[[[0.          0.          0.00283302 ... 0.00121139 0.00177172 0.00240969]
  [0.00439186 0.00189388 0.          ... 0.00251231 0.          0.          ]]
 (2, 1921)]

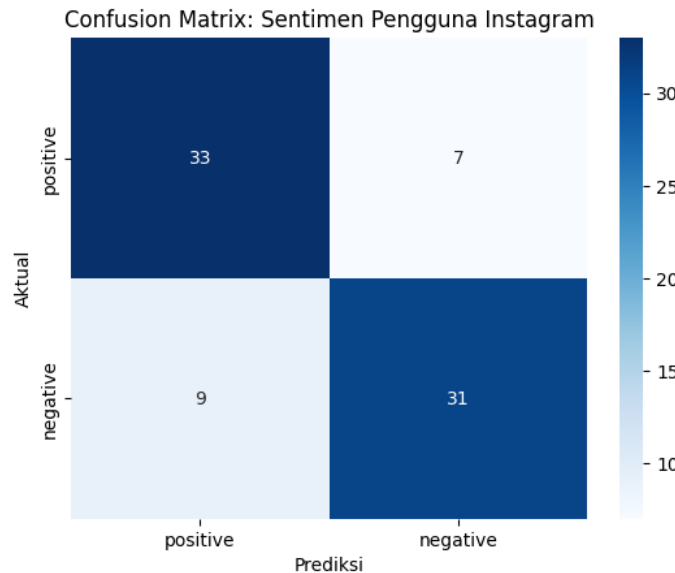
print(model.var_)
print(model.var_.shape)

[[[6.80875057e-12 6.80875057e-12 1.27613556e-03 ... 2.33325241e-04
  4.99098028e-04 9.23249378e-04]
 [1.55523721e-03 5.70300278e-04 6.80875057e-12 ... 1.00356331e-03
  6.80875057e-12 6.80875057e-12]]]
 (2, 1921)]
```

Gambar 3 Model Gaussian Naive Bayes

3.2 Analisis Sistem

tahapan analisis sistem bertujuan untuk memaparkan hasil model *Gaussian Naive Bayes* dalam menganalisis data komentar instagram. Analisis sistem ini memaparkan nilai *accuracy*, *precision*, *recall*, dan *F1-score* dengan menggunakan metode *confusion matrix* untuk mengevaluasi kinerja model klasifikasi *Gaussian Naive Bayes*.



Gambar 4 Confusion Matrix
Tabel 6 Classification Report

	Precision	Recall	F1-Score
Negative	0.82	0.78	0.79
Positive	0.79	0.82	0.80
Accuracy			0.80

Berdasarkan tabel 6 pada pengujian hasil klasifikasi *Gaussian Naive Bayes* baik ditunjukkan dengan nilai akurasi 80%. Pada pengujian yang telah dilakukan menunjukkan bahwa untuk kelas negative menghasilkan nilai precision 82%, Recall 78%, F1-Score 89%. sedangkan untuk kelas positive menghasilkan nilai precision 79%, Recall 82%, F1-Score 80%.

Dalam table 6 menunjukkan Precision untuk kelas negatif (82%) lebih tinggi dibandingkan kelas positif yang hanya (79%) menunjukkan bahwa model lebih baik dalam mengidentifikasi True Negative. Pada bagian Recall menunjukkan kelas positif (82%) lebih tinggi dibandingkan dengan kelas negatif (78%) menunjukkan bahwa model lebih baik dalam mengidentifikasi True Positive. Pada bagian F1-score menunjukkan kelas negative (89%) lebih tinggi dibandingkan kelas positif (80%) menunjukkan rata-rata harmonis yang lebih baik antara Precision dan recall di kelas negatif.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah diselesaikan, berikut adalah kesimpulan pada penelitian Implementasi Algoritma Klasifikasi Naive Bayes Untuk Analisis Sentimen Teks Berbahasa Indonesia Pada Komentar Instagram berhasil diimplementasikan dan memberikan hasil yang baik dalam menganalisis sentimen. Penelitian ini menggunakan dataset komentar instagram dan algoritma Naive bayes mendapatkan nilai akurasi 80%. untuk kelas negative menghasilkan nilai precision 82%, Recall 78%, F1-Score 79%. sedangkan untuk kelas positive menghasilkan nilai precision 79%, Recall 82%, F1-Score 80%. F1-score menunjukkan kelas

negative (89%) lebih tinggi dibandingkan kelas positif (80%) menunjukkan rata-rata harmonis yang lebih baik antara Precision dan recall di kelas negative

REFERENCES

- Abdul Mohaimin Rahat, A. K. (2019). Comparison of Naive Bayes and SVM Algorithm. *Proceedings of the SMART-2019*, 266-270.
- Alnaz, F. S. (2021). Analisis Emosi Melalui Media Sosial Twitter Dengan Menggunakan Metode Naïve Bayes dan Perbandingan Fitur N-gram dan TF-IDF. *Proceedings of Engineering*, 8(5).
- Angelina Puput Giovani, A. T. (2020). ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER. *Jurnal TEKNOINFO*, 116-124.
- Aripin, W. A. (2020). Sequential Model for Mapping Compound Emotions in Indonesian Sentences. *Journal of Applied Intelligent System*, Vol.5 No.1, 32-46.
- Candra, R. M. (2020). Klasifikasi Komentar Bullying pada Instagram Menggunakan Metode K-Nearest Neighbor. *IT Journal Research and Development*, 5(1), 45–52.
- Junianto, E. &. (2020). Penerapan Metode Naïve Bayes Classifier Untuk Mendeteksi Emosi Pada Komentar Media Sosial. *Jurnal Responsif: Riset Sains dan Informatika*, 2(1), 1-8.
- M Indra Halim Arsyah Dwi Akbari, A. N. (2017). Analisis Sentimen Menggunakan Metode Learning Vector Quantization. *e-Proceeding of Engineering*, 2283.
- M. Krommyda, A. R. (2020). Emotion detection in Twitter posts: a rule-based algorithm for annotated data acquisition. *International Conference on Computational Science and Computational Intelligence*, 257-262.
- Muhammad Abbas, K. A. (2019). Multinomial Naive Bayes Classification Model for Sentiment . *International Journal of Computer Science and Network Security*, 62-67.
- Muhammad Zubair Asghar, A. L.-R.-A. (2022). A Deep Neural Network Model for the Detection and Classification of Emotions from Textual Content. *Complexity*, 12.
- Mustafa, M. S. (2018). Leksikon untuk Deteksi Emosi dari Teks Bahasa Indonesia. *Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier*, 4(2), 151-162.
- Putu, N. L. (2021). Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation. *Rekayasa Sistem Dan Teknologi Informasi*, 5(1), 123 - 131.
- Ramadhani, S. H. (2022). Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naïve Bayes dan K-NN. *Jurnal Teknologi Informasi Dan Komunikasi*, 27(1), 526–534.
- Rohman, A. U. (2019). Deteksi Kondisi Emosi pada Media Sosial Menggunakan Pendekatan Leksikon dan Natural Language Processing. *Eksplora Informatika*, 9(1), 70-76.
- Rona Nisa Sofia Amriza, &. D. (2021). Komparasi Metode Machine Learning dan Deep Learning untuk Deteksi Emosi pada Text di Sosial Media. *Jurnal Penelitian Ilmu Dan Teknologi Komputer*, 13(2), 130–139.
- Sepyan Purnama Kristanto, J. A. (2019). Naive Bayes Classifier on Twitter Sentiment . *International Conference of Computer and Informatics Engineering*, 24-28.
- Z. Rajabi, A. S. (2020). A Multi-channel BiLSTM-CNN Model for Multilabel Emotion Classification of Informal Text. *International Conference on Semantic Computing*, 303-306.