Jurnal Riset Multidisiplin Edukasi
https://journal.hasbaedukasi.co.id/index.php/jurmie
Halaman : 717 - 729

Analisis Sentimen Masyarakat Terkait Gaji Pns Rendah Sebagai Faktor Korupsi Memakai Metode Naïve Bayes Classifier (Studi Kasus: Akun Kumparancom Di Instagram)

Muhammad Iqbal Maulana[1], Bambang Santoso[2]
Ilmu Komputer, Teknik Informatika, Universitas Pamulang [1,2]
Email: iqbalabala763@gmail.com

Informasi

Volume : 3
Nomor : 7
Bulan : Juli
Tahun : 2026
E-ISSN : 3062-9624

Abstract

Low salaries of Indonesian civil servants (PNS) are often associated with corruption and widely discussed on social media. This study aims to analyze public sentiment regarding this issue using the Naïve Bayes Classifier method. A total of 1,165 comments were collected from the Kumparancom Instagram account between October 2023 and February 2024 through a crawling process. The data were preprocessed using cleaning, case folding, tokenizing, normalization, stopword removal, and stemming, then classified into positive and negative sentiments. The results indicate that public sentiment is predominantly negative toward the assumption that low civil servant salaries are the main cause of corruption. Model evaluation using a confusion matrix shows that the Naïve Bayes Classifier achieved the highest accuracy of 95% with a 70% training and 30% testing data split. These findings confirm that the proposed method is effective for sentiment analysis of social media text.

Keyword: *Sentiment Analysis, Naïve Bayes Classifier, Instagram, Civil Servant Salary, Corruption.*

Abstrak

Rendahnya gaji Pegawai Negeri Sipil (PNS) sering dikaitkan dengan meningkatnya praktik korupsi di Indonesia dan menjadi perbincangan di media sosial. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap isu tersebut menggunakan metode Naïve Bayes Classifier. Data diperoleh dari 1.165 komentar netizen pada akun Instagram Kumparancom periode Oktober 2023 hingga Februari 2024 melalui proses crawling. Data kemudian melalui tahapan preprocessing yang meliputi cleaning, case folding, tokenizing, normalisasi, stopword removal, dan stemming, sebelum diklasifikasikan ke dalam sentimen positif dan negatif. Hasil penelitian menunjukkan bahwa sentimen masyarakat didominasi oleh sentimen negatif terhadap anggapan rendahnya gaji PNS sebagai faktor utama korupsi. Evaluasi model menggunakan confusion matrix menunjukkan bahwa metode Naïve Bayes Classifier menghasilkan akurasi tertinggi sebesar 95% pada pembagian data latih dan data uji 70%:30%. Hasil ini menunjukkan bahwa metode tersebut efektif untuk analisis sentimen berbasis teks pada media sosial.

Kata Kunci: *Analisis Sentimen, Naïve Bayes Classifier, Instagram, Gaji PNS, Korupsi.*

A. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah menjadikan media sosial sebagai sarana utama masyarakat dalam menyampaikan opini terhadap berbagai isu sosial dan pemerintahan. Media sosial memungkinkan interaksi dua arah yang cepat dan terbuka, sehingga menghasilkan volume data teks yang besar dan beragam. Instagram merupakan salah satu platform media sosial yang banyak digunakan di Indonesia dan sering dimanfaatkan sebagai ruang diskusi publik terhadap isu-isu aktual.

Salah satu isu yang kerap menjadi perbincangan adalah rendahnya gaji Pegawai Negeri Sipil (PNS) yang dianggap memiliki keterkaitan dengan praktik korupsi di Indonesia. Isu ini memunculkan berbagai pandangan di masyarakat, baik yang mendukung maupun menolak anggapan bahwa rendahnya gaji PNS merupakan faktor utama terjadinya korupsi. Banyaknya opini yang muncul di media sosial menjadikan platform tersebut sebagai sumber data yang potensial untuk dianalisis guna mengetahui kecenderungan sentimen masyarakat.

Analisis sentimen merupakan salah satu penerapan *Natural Language Processing* (NLP) yang digunakan untuk mengklasifikasikan opini atau pendapat dalam bentuk teks ke dalam kategori sentimen tertentu. Metode *Naïve Bayes Classifier* dipilih dalam penelitian ini karena memiliki algoritma yang sederhana, efisien, serta mampu memberikan hasil klasifikasi yang baik pada data teks. Kontribusi penelitian ini terletak pada penerapan analisis sentimen berbasis Instagram untuk mengkaji persepsi masyarakat terhadap isu gaji PNS dan korupsi, serta evaluasi kinerja metode *Naïve Bayes Classifier* pada data komentar berbahasa Indonesia.

B. METODE PENELITIAN

2.1. Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis sentimen berbasis machine learning. Tujuan penelitian adalah mengklasifikasikan opini masyarakat terhadap isu gaji PNS rendah sebagai faktor korupsi di Indonesia ke dalam sentimen positif dan negatif menggunakan metode Naïve Bayes Classifier.

Data yang digunakan merupakan data sekunder yang diperoleh dari komentar pengguna Instagram pada akun Kumparancom, yang membahas isu terkait. Analisis dilakukan melalui tahapan pengumpulan data, preprocessing, pelabelan sentimen, klasifikasi, dan evaluasi model.

2.2. Sumber dan Pengumpulan Data

Data dikumpulkan melalui teknik crawling menggunakan ekstensi Instant Data Scraper pada browser Google Chrome. Proses crawling dilakukan pada kolom komentar salah satu

unggahan Instagram Kumparancom yang membahas isu gaji PNS dan korupsi. Data yang berhasil dikumpulkan berjumlah 1.165 komentar dan disimpan dalam format XLSX untuk selanjutnya diproses.

Tabel 1. Deskripsi Dataset

Keterangan	Nilai
Sumber Data	Instagram Kumparancom
Teknik Pengambilan	Crawling
Jumlah Data	1.165 komentar
Periode Data	November 2023
Format Data	XLSX

2.3. Perangkat Penelitian

Tabel 2. Perangkat Penelitian

Jenis Perangkat	Nama/ Spesifikasi
Perangkat Lunak	Jupyter Notebook
Perangkat Lunak	Python 3
Perangkat Lunak	Instant Data Scraper versi 1.0.8
Sistem Operasi	Windows 11 64-bit
Perangkat Keras	Laptop HP Pavilion Gaming
Prosesor	Intel Core i7-10750H
RAM	16 GB

2.4. Pelabelan Sentimen

Pelabelan data dilakukan secara manual untuk menentukan polaritas sentimen pada setiap komentar. Data diklasifikasikan ke dalam dua kelas sentimen, yaitu positif dan negatif, berdasarkan makna dan konteks kalimat.

2.5. Preprocessing Data

2.5.1 Cleaning Data

Cleaning Data membersihkan komentar dari tanda baca, mention, hastag, emoticon dan karakter lainnya yang kurang penting yang bertujuan untuk mengurangi noise.

2.5.2 Case Folding

Case folding merupakan tahapan preprocessing data dimana karakter yang ada pada komentar akan diubah atau dikonversi menjadi huruf kecil.

2.5.3 Tokenize

Tokenize untuk melakukan pemisahan kata dalam suatu kalimat dengan tujuan untuk proses analisis teks lebih lanjut. Tiap kata diberikan tanda petik dan koma Pada proses ini juga

dilakukan proses menghilangkan tanda baca seperti titik, koma, spasi dan karakter angka.

2.5.4 Normalize

Normalisasi adalah metode yang dipakai untuk menggabungkan data dari beragam entitas dalam suatu relasi sehingga terbentuk struktur yang teratur tanpa duplikasi data. Tujuannya adalah untuk mengurangi data yang berulang, menjamin penempatan data yang tepat, menghilangkan data yang tak cocok, dan memastikan data yang diubah sesuai harapan.

2.5.5 Stopword Removal

Stopword Removal Langkah Penghapusan Stopword adalah proses yang dilakukan untuk mengeliminasi kata-kata yang tidak memiliki arti sama sekali. Contoh kata-kata nya adalah "dan", "juga", "di", "ke", "dari", "itu", "saya", "kamu", "mereka".

2.5.6 Stemming

Stemming adalah menghilangkan awalan atau akhiran kata sehingga hanya menyisakan kata dasarnya.

2.6. Pembuatan Model Klasifikasi

Dataset yang telah melalui tahap pelabelan dan preprocessing digunakan untuk membangun model klasifikasi sentimen. Data dibagi menjadi data training dan data testing untuk proses pelatihan dan pengujian model. Metode klasifikasi yang digunakan adalah Naïve Bayes Classifier yang diimplementasikan menggunakan library scikit-learn pada bahasa pemrograman Python.

C. HASIL DAN PEMBAHASAN

3.1. Hasil Klasifikasi Sentimen

Berdasarkan hasil klasifikasi sentimen menggunakan metode *Naïve Bayes Classifier*, diperoleh gambaran bahwa sentimen masyarakat terhadap isu rendahnya gaji Pegawai Negeri Sipil (PNS) sebagai faktor terjadinya korupsi didominasi oleh sentimen negatif. Hal ini menunjukkan bahwa mayoritas pengguna Instagram yang memberikan komentar pada akun Kumparancom cenderung tidak sepenuhnya menyetujui anggapan bahwa rendahnya gaji PNS merupakan penyebab utama praktik korupsi.

Komentar dengan sentimen negatif umumnya mengandung opini yang menekankan bahwa korupsi lebih dipengaruhi oleh faktor integritas individu, mentalitas, serta lemahnya sistem pengawasan, bukan semata-mata karena faktor ekonomi. Sementara itu, komentar dengan sentimen positif umumnya menyatakan bahwa gaji yang rendah dapat menjadi salah satu pemicu korupsi, terutama pada tingkat jabatan tertentu. Namun, jumlah sentimen positif

lebih sedikit dibandingkan sentimen negatif, sehingga dapat disimpulkan bahwa persepsi publik cenderung kritis terhadap narasi yang menyederhanakan penyebab korupsi hanya dari sisi penghasilan PNS.

3.2. Evaluasi Kinerja Model Klasifikasi

Evaluasi kinerja model klasifikasi dilakukan menggunakan *confusion matrix* dengan metrik pengukuran meliputi *accuracy, precision, recall,* dan *F1-score*. Proses evaluasi dilakukan pada beberapa skenario pembagian data training dan testing, dan hasil terbaik diperoleh pada pembagian 70% data training dan 30% data testing.

Pada skenario tersebut, model *Naïve Bayes Classifier* berhasil mencapai akurasi sebesar 95%, yang menunjukkan bahwa model mampu mengklasifikasikan sentimen komentar dengan tingkat ketepatan yang sangat tinggi. Nilai precision dan recall yang seimbang menunjukkan bahwa model tidak hanya mampu mengidentifikasi sentimen negatif dengan baik, tetapi juga cukup akurat dalam mengenali sentimen positif. Skor F1 yang tinggi mengindikasikan bahwa model memiliki performa yang stabil dan konsisten dalam mengklasifikasikan data teks berbahasa Indonesia.

Tingginya nilai akurasi ini dipengaruhi oleh proses preprocessing yang dilakukan secara menyeluruh, mulai dari *cleaning data, case folding, tokenizing,* normalisasi, *stopword removal,* hingga *stemming*. Tahapan tersebut terbukti mampu mengurangi noise dan meningkatkan kualitas fitur teks yang digunakan dalam proses klasifikasi.

3.3. Pembahasan

Hasil analisis sentimen menunjukkan bahwa dominasi sentimen negatif mencerminkan pandangan kritis masyarakat terhadap isu gaji PNS dan korupsi. Masyarakat menilai bahwa korupsi merupakan permasalahan yang kompleks dan tidak dapat dijelaskan hanya melalui satu faktor, seperti rendahnya gaji. Faktor moral, budaya organisasi, serta efektivitas penegakan hukum dinilai memiliki peran yang lebih signifikan.

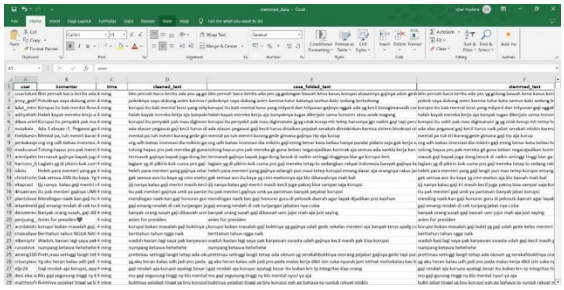
Dari sisi metodologi, penerapan *Naïve Bayes Classifier* terbukti efektif dalam melakukan analisis sentimen pada data komentar Instagram. Metode ini mampu menangani data teks dalam jumlah besar dengan proses komputasi yang relatif sederhana dan efisien. Hasil penelitian ini sejalan dengan penelitian-penelitian sebelumnya yang menyatakan bahwa *Naïve Bayes Classifier* memiliki performa yang baik dalam tugas klasifikasi teks, khususnya pada analisis sentimen media sosial.

Dengan demikian, penelitian ini tidak hanya memberikan gambaran mengenai persepsi masyarakat terhadap isu gaji PNS dan korupsi, tetapi juga menunjukkan bahwa metode *Naïve*

Bayes Classifier dapat dijadikan sebagai pendekatan yang andal untuk analisis sentimen berbasis media sosial berbahasa Indonesia.

4. IMPLEMENTASI

Dalam studi ini, penelitian menerapkan temuan dari penelitian sebelumnya terhadap kumpulan data berisi 966 komentar tentang gaji pns rendah pemicu korupsi di indonesia.



Gambar 1. *Dataset telah melalui Preprocessing Data*

Dalam penerapannya, penelitian ini menggunakan metode *Naïve bayes classifier* dan kemudian mengevaluasi performanya dengan menggunakan *confusion matrix* yang memiliki lima kategori untuk mengukur kinerja model tersebut. Lima kategori tersebut diantaranya adalah *accuracy, recall, precision, f1-score* dan *macro avg.*

4.1. Pengujian

Pada tahap Pengujian, peneliti menerapkan Metode *Naïve Bayes Classifier* untuk mengklasifikasikan sentimen. Dalam setiap data dibagi menjadi 3 bagian, yaitu untuk kelas pertama data latih 70% dan data uji 30%, lalu data kedua data latih 60% dan data uji 40%, selanjutnya kelas ke ketiga data latih 50% dan data uji 50%. Setiap data yang dipisah, kemudia akan diprediksi menggunakan library NLTK berdasarkan metode *Naïve bayes*. Berdasarkan hasil evaluasi nanti, dapat dipilih salah satu kelas yang memiliki performa yang paling baik untuk digunakan dalam analisis atau pengambilan keputusan lebih lanjut.

4.1.1 Pengujian Data Kelas Satu

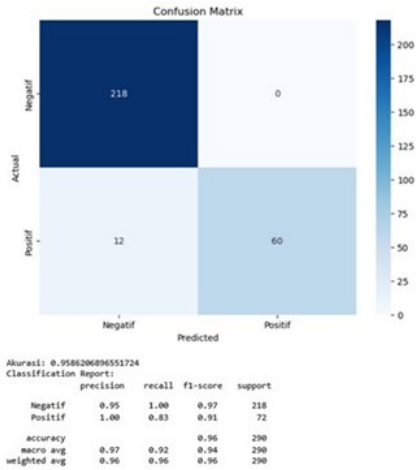
Dalam proses pembagian data untuk keperluan pelatihan dan pengujian kelas pertama, yaitu 70% dari keseluruhan data, yakni 350 untuk data training, sedangkan 30% sisanya, yakni 151 untuk data uji. Fokus utama penelitian ini adalah untuk mencapai tingkat akurasi yang optimal. Dalam proses klasifikasi yang telah diuraikan, ada berbagai skenario yang dirancang untuk mengukur da memperoleh nilai *accuracy, precision, recall*, serta *F1 Score*. Jumlah sentimen dalam data yang telah dipisahkan menjadi data training dan data uji adalah sebagai berikut:

Tabel 1. *Hasil & Pembagian Data Latih – Data Uji Kelas Satu*

	Positif	**Negatif**

Data Latih	113	563
Data Uji	72	218
Total Data Positif	185	
Total Data Negatif	781	
Total Data	966	

Setelah melakukan pengujian klasifikasi Naïve Bayes menggunakan data uji, langkah berikutnya adalah menganalisis hasil klasifikasi dengan menggunakan *Confusion Matrix* dalam *source code Python* yang dieksekusi dalam lingkungan *Jupyter Notebook*. Confusion Matrix adalah bentuk tabel yang menunjukkan hasil klasifikasi yang telah diperoleh, termasuk informasi tentang *true positive, true negative, false positive,* dan *false negative*.



Gambar 2. *Confusion Matrix Data Uji Kelas 1*

Berdasarkan *confusion matrix* diatas menghasilkan beberapa nilai hasil kinerja suatu model klasifikasi, diantaranya nilai *accuracy*, precesion, recall dan skor f1. Berikut adalah perhitungan detail dari hasil kinerja suatu model klasifikasi:

a. *Accuracy*

Melihat seberapa akurat model dapat mengklasifikasikan dengan benar tingkat kedekatan.

$$\textbf{Accuracy} = \frac{TP+TN}{TP+TN+FN+FP} \textbf{ x 100\%} = \frac{60+218}{60+218+0+12} = \frac{278}{290} \textbf{= 95 x 100\% = 95\%}$$

b. *Precision*

Memprediksi benar positif dibandingkan dengan keseluruhan hasil yanf diprediksi positif.

$$\textbf{Precision} = \frac{TP}{TP+FN} \textbf{ x 100\%} = \frac{60}{60+0} = \frac{60}{60} \textbf{= 1 x 100 \% = 1\%}$$

c. *Recall*

Memprediksi benar positif dibandingkan dengan keseluruhan data yang benar positif

keberhasilan model dalam menemukan kembali sebuah informasi.

Recall = $\frac{TP}{TP+FN}$ x 100% = $\frac{60}{60+12}$ = $\frac{60}{72}$ = 0,83 x 100% = 83%

d. Skor F1

Matrix gabungan yang menggabungkan presisi dan *recall* dalam suatu nilai tunggal. Ini berguna ketika ingin mengukur keseimbangan antara presisi dan recall.

Skor F1 = $\frac{2\ x\ Presisi\ x\ Recal}{Presisi+Recall}$ = $\frac{2\ x\ 1{,}00\ x\ 0{,}83}{1{,}00+0{,}83}$ =$\frac{1{,}66}{1{,}83}$ = 0,91 x 100% = 91%

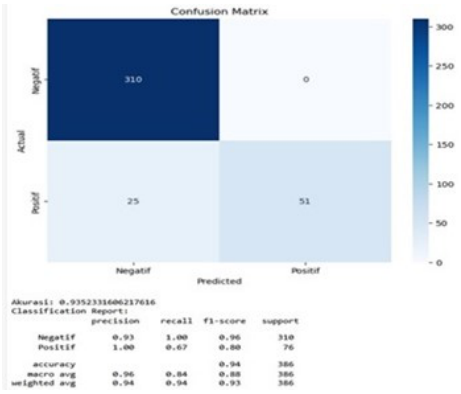
4.1.2 Pengujian Data Kelas Dua

Dalam proses pembagian data untuk keperluan pelatihan dan pengujian kelas kedua, yaitu 60% dari keseluruhan data, yakni 306 untuk data training, sedangkan 40% sisanya, yakni 205 untuk data uji. Fokus utama penelitian ini adalah untuk mencapai tingkat akurasi yang optimal. Dalam proses klasifikasi yang telah diuraikan, ada berbagai skenario yang dirancang untuk mengukur dan memperoleh nilai *accuracy, precision, recall,* serta *F1 Score*. Jumlah sentimen dalam data yang telah dipisahkan menjadi data training dan data uji adalah sebagai berikut:

Tabel 2. *Hasil & Pembagian Data Latih – Data Uji Kelas Dua*

	Positif	**Negatif**
Data Latih	96	484
Data Uji	76	310
Total Data Positif	**172**	
Total Data Negatif	**794**	
Total Data	**966**	

Setelah melakukan pengujian klasifikasi Naïve Bayes menggunakan data uji, langkah berikutnya adalah menganalisis hasil klasifikasi dengan menggunakan *Confusion Matrix* dalam *source code Python* yang dieksekusi dalam lingkungan *Jupyter Notebook*. *Confusion Matrix* adalah bentuk tabel yang menunjukkan hasil klasifikasi yang telah diperoleh, termasuk informasi tentang *true positive, true negative, false positive,* dan *false negative*.



Gambar 3. *Confusion Matrix Data Uji Kelas 2*

Berdasarkan *confusion matrix* diatas menghasilkan beberapa nilai hasil kinerja suatu model klasifikasi, diantaranya nilai *accuracy*, *precesion, recall* dan *skor f1*. Berikut adalah perhitungan detail dari hasil kinerja suatu model klasifikasi:

a. *Accuracy*

Melihat seberapa akurat model dapat mengklasifikasikan dengan benar tingkat kedekatan.

$$\textbf{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \textbf{ x 100} = \frac{51+310}{51+310+25} = \frac{361}{386} \textbf{ = 0,93 x 100\% = 93\%}$$

b. *Precision*

Memprediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif.

$$\textbf{Precision} = \frac{TP}{TP+FP} \textbf{ x 100\% } = \frac{51}{51+0} = \frac{51}{51} \textbf{ = 1,00 x 100\% = 1\%}$$

c. *Recall*

Memprediksi benar positif dibandingkan dengan keseluruhan data yang benar positif keberhasilan model dalam menemukan kembali sebuah informasi.

$$\textbf{Recall} = \frac{TP}{TP+FN} \textbf{ x 100\% } = \frac{51}{51+25} = \frac{51}{76} \textbf{ = 0,67 x 100\% = 67\%}$$

d. Skor F1

Matrix gabungan yang menggabungkan presisi dan *recall* dalam suatu nilai tunggal. Ini berguna ketika ingin mengukur keseimbangan antara presisi dan recall.

$$\textbf{Skor F1} = \frac{2\ x\ Precision\ x\ Recall}{Precision+Recall} = \frac{2x\ 1,00\ x\ 0,67}{1,00+0,67} = \frac{1,34}{1,67} \textbf{ = 0,80 x 100\% = 80\%}$$

4.1.3 Pengujian Data Kelas Tiga

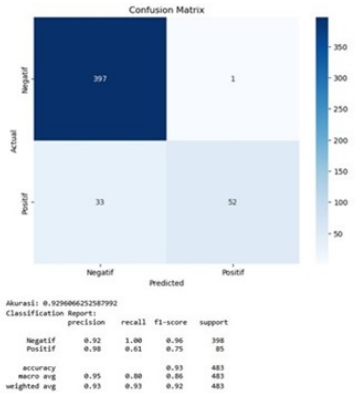
Dalam proses pembagian data untuk keperluan pelatihan dan pengujian kelas ketiga, yaitu 50% dari keseluruhan data, yakni 483 untuk data training, sedangkan 50% sisanya, yakni 483 untuk data uji. Fokus utama penelitian ini adalah untuk mencapai tingkat akurasi yang optimal. Dalam proses klasifikasi yang telah diuraikan, ada berbagai skenario yang dirancang untuk mengukur da memperoleh nilai accuracy, precision, recall, serta F1 Score. Jumlah sentimen dalam data yang telah dipisahkan menjadi data training dan data uji adalah sebagai berikut:

Tabel 3. *Hasil & Pembagian Data Latih – Data Uji Kelas Tiga*

	Positif	**Negatif**
Data Latih	**85**	**398**
Data Uji	**85**	**398**
Total Data Positif	**483**	
Total Data Negatif	**483**	

Total	**966**

Setelah melakukan pengujian klasifikasi Naïve Bayes menggunakan data uji, langkah berikutnya adalah menganalisis hasil klasifikasi dengan menggunakan *Confusion Matrix* dalam *source code Python* yang dieksekusi dalam lingkungan *Jupyter Notebook. Confusion Matrix* adalah bentuk tabel yang menunjukkan hasil klasifikasi yang telah diperoleh, termasuk informasi *tentang true positive, true negative, false positive,* dan *false negative*.



Gambar 4. *Confusion Matrix Data Uji Kelas 3*

Berdasarkan *confusion matrix* diatas menghasilkan beberapa nilai hasil kinerja suatu model klasifikasi, diantaranya nilai *accuracy*, *precesion, recall* dan *skor f1*. Berikut adalah perhitungan detail dari hasil kinerja suatu model klasifikasi:

a. *Accuracy*

Melihat seberapa akurat model dapat mengklasifikasikan dengan benar tingkat kedekatan.

$$\textbf{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \textbf{ x 100} = \frac{52+397}{52+397+1+33} = \frac{449}{483} \textbf{ = 0,92 x 100\% = 92\%}$$

b. *Precision*

Memprediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif.

$$\textbf{Precision} = \frac{TP}{TP+FP} \textbf{ x 100\%} = \frac{52}{53+0} = \frac{52}{53} \textbf{ = 0,98 x 100\% = 98\%}$$

c. *Recall*

Memprediksi benar positif dibandingkan dengan keseluruhan data yang benar positif keberhasilan model dalam menemukan kembali sebuah informasi.

$$\textbf{Recall} = \frac{TP}{TP+FN} \textbf{ x 100\%} = \frac{52}{85} \textbf{ = 0,61 x 100\% = 61\%}$$

d. Skor F1

Matrix gabungan yang menggabungkan presisi dan *recall* dalam suatu nilai

tunggal. Ini berguna ketika ingin mengukur keseimbangan antara presisi dan *recall*.

$$\textbf{Skor F1} = \frac{2\ x\ Precision\ x\ Recall}{Precision+Recall} = \frac{2\ x\ 0{,}98\ x\ 0{,}61}{0{,}98+0{,}61} = \frac{1{,}19}{1{,}59} = 0{,}75 \text{ x } 100\% = 75\%$$

4.2 Pembahasan Implementasi dan Pengujian

4.2.1 Pembahasan Implementasi Sistem

Berdasarkan hasil pelaksanaan sistem yang telah diimplementasikan, kesimpulan yang dapat ditarik dari implementasi tersebut adalah:

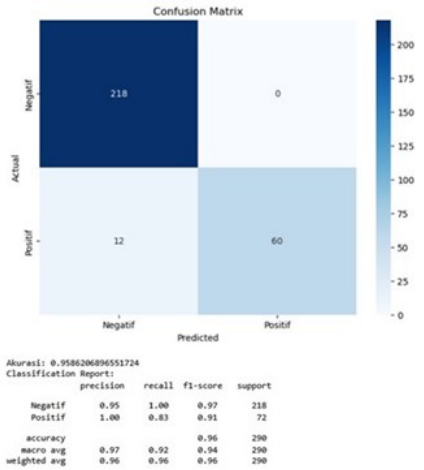
a. Kebutuhan perangkat keras dan perangkat lunak memadai sehingga mempermudah Preprocessing data mulai dari *cleaning data, Case folding, Tokinize, Normalized, Stopword Removal* dan *Stemming*.
b. Dataset komentar Instagram yang telah melalui proses pembersihan dari kata-kata tak bermakna dan tanda baca dapat digunakan pada tahap berikutnya, yaitu tahap pengujian.

4.2.2 Pembahasan Pengujian Klasifikasi *Naïve Bayes Classifier*

Dari hasil pengujian tiga kelas klasifikasi *Naïve Bayes* yang telah dilakukan, dapat disimpulkan dalam bentuk tabel *confusion matrix* sebagai berikut:

a. Data Kelas Satu

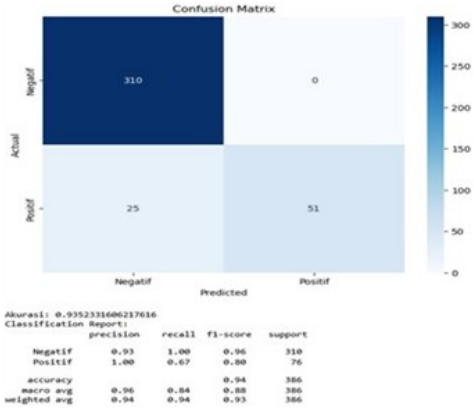
Untuk data uji 30% dengan total data 290. Terdapat komentar 72 positif dan 218 negatif dengan nilai *accuracy* sebesar 0.95, presisi sebesar 1.00, *recall* sebesar 0.82, dan nilai *F1 Score* 83.



Gambar 5. *Hasil Perhitungan naïve Bayes Classifier Kelas Satu*

b. Data Kelas Dua

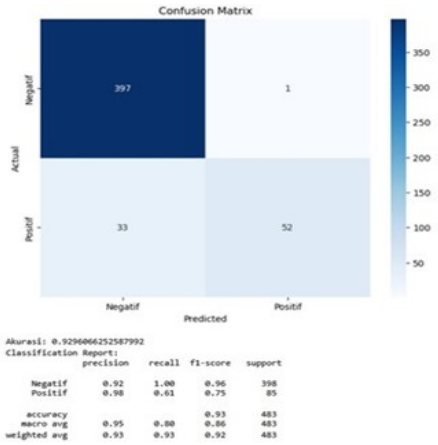
Untuk daja uji 40% dengan total data 386. Terdapat komentar 96 positif dan 484 negatif dengan nilai *accuracy* sebesar 0.95, presisi sebesar 1.00, *recall* sebesar 0.67, dan nilai *F1 Score* 0.80.



Gambar 6. *Hasil Perhitungan Naïve Bayes Classifier Kelas Dua*

c. Data Kelas Tiga

Untuk data uji 50% dengan total data 483. Terdapat komentar 85 positif dan 398 negatif dengan nilai *accuracy* sebesar 0.92, presisi sebesar 0.98, *recall* sebesar 0.61, dan nilai *F1 Score* 0.75.



Gambar 7. *Hasil Perhitungan Naïve Bayes Classifier Kelas Tiga*

Dari hasil klasifikasi dengan tiga kelas data training dengan data testing dapat dilihat pada rangkuman tabel berikut:

Tabel 4. *Hasil Komposisi Akurasi Dari Tiga Kelas*

Komposisi	Akurasi	Presisi	Recall	F1-Score
70%:30%	0,95	1	0,82	0,83
60%:40%	0,93	1	0,67	0,80
50%:50%	0,92	1	0,62	0,75

Dari tabel diatas akurasi tertinggi diperoleh dengan komposisi jumlah data training: Jumlah data testing yaitu sebesar 70%:30%. Dengan demeikian akurasi tertinggi metode *Naïve Bayes Classifier* adalah sebesar 95%.

D. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa sentimen masyarakat di Instagram terhadap anggapan rendahnya gaji PNS sebagai faktor utama korupsi di Indonesia didominasi oleh sentimen negatif. Hal ini menunjukkan bahwa masyarakat cenderung menilai faktor lain, seperti integritas dan sistem pengawasan, memiliki peran yang lebih besar dalam terjadinya korupsi.

Metode *Naïve Bayes Classifier* menunjukkan kinerja yang baik dalam analisis sentimen dengan akurasi tertinggi sebesar 95% pada pembagian data latih dan data uji 70%:30%. Dengan demikian, metode ini dapat digunakan secara efektif untuk analisis sentimen berbasis teks pada media sosial. Penelitian ini diharapkan dapat menjadi referensi bagi pengembangan analisis sentimen pada isu-isu sosial dan pemerintahan di masa mendatang.

E. DAFTAR PUSTAKA

Di, I., & Garden, G. (2024). RELASI : Jurnal Penelitian Komunikasi , RELASI : Jurnal Penelitian Komunikasi ,. 04(03), 11–18.

Ikhsan Yusuf, M., Ransi, N., Arman, A., Tenriawaru, A., Saidi, L. O., & La Surimi, L. S. (2022). Analisis Sentimen Komentar Netizen Instagram Terhadap Racism Di Sepak Bola Indonesia Dengan Metode Naive Bayes. Jurnal Matematika Komputasi Dan Statistika, 2(3), 136–143. https://doi.org/10.33772/jmks.v2i3.10

Rahayu, I. P., Fauzi, A., & Indra, J. (2022). Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine. 4, 296–301. https://doi.org/10.30865/json.v4i2.5381

Sihombing, D. O. (2022). Implementasi Natural Language Processing (NLP) dan Algoritma Cosine Similarity dalam Penilaian Ujian Esai Otomatis. 4, 396–406. https://doi.org/10.30865/json.v4i2.5374

Wiryany, D., Natasha, S., & Kurniawan, R. (2022). KOMUNIKASI TERHADAP PERUBAHAN SISTEM. 8(November), 242–252.

Yasin, R. Al, Roro, R., Annisa, K., & Salsabil, S. (2022). PENGARUH SOSIAL MEDIA TERHADAP KESEHATAN MENTAL DAN FISIK REMAJA : A SYSTEMATIC REVIEW. 3, 82–90.