

**ANALISIS SENTIMEN PENGGUNA INSTAGRAM TERHADAP
STATEMENT MENTERI KEUANGAN TENTANG
“KEBIJAKAN GAJI GURU DAN DOSEN”
MENGUNAKAN NAÏVE BAYES**

M. Imam Santoso¹, Nia Mardiah², Asha Sembiring³
Universitas Satya Terra Bhinneka, Medan

e-mail: ¹mhdimamsantoso@satyaterrabhinneka.ac.id,

²niamardiah@satyaterrabhinneka.ac.id, ³ashasembiring@satyaterrabhinneka.ac.id

Abstract: *The Minister of Finance's statement regarding the teacher and lecturer salary policy sparked mixed public reactions, particularly on social media. This study aims to analyze the sentiment of the Indonesian public, particularly Instagram users, regarding this statement using the Naïve Bayes algorithm. Data was collected through scraping using Instant Data Scraper in August 2025, totaling 939 data points. Preprocessing steps included cleansing, tokenizing, stopword removal, and stemming. The data were then classified into two sentiment categories: positive and negative. The results showed that the majority of Instagram user sentiment tended to be negative (859 data points, representing 91.5%), while positive sentiment accounted for 89 data points, representing 9.5%. The Naïve Bayes model achieved an accuracy of 0.87 in classifying public opinion. These findings indicate that the Naïve Bayes algorithm is effective in analyzing public opinion on sensitive issues on social media. Furthermore, the results of this study can serve as a reference for the government and policymakers in understanding public perception and formulating more appropriate communication strategies related to education policy.*

Keyword: *Sentiment Analysis, Naïve Bayes, Minister of Finance, Instagram, Text Analysis.*

Abstrak: Pernyataan Menteri Keuangan tentang kebijakan gaji guru dan dosen memicu beragam reaksi publik, khususnya di media sosial. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia, terutama pengguna Instagram, terhadap pernyataan tersebut dengan menggunakan algoritma Naïve Bayes. Data dikumpulkan melalui proses scraping menggunakan Instant Data Scraper pada Agustus 2025 dengan jumlah 939 data. Kemudian dilakukan tahap praproses meliputi cleansing, tokenizing, stopword removal, dan stemming. Selanjutnya, data diklasifikasikan ke dalam dua kategori sentimen, yaitu positif dan negatif. Hasil penelitian menunjukkan bahwa mayoritas sentimen pengguna Instagram cenderung Negatif (859 data dengan persentase 91,5%), sedangkan sentimen positif berjumlah 89 data dengan persentase 9,5%. Model Naïve Bayes mencapai tingkat akurasi sebesar 0,87 dalam mengklasifikasikan opini publik. Temuan ini mengindikasikan bahwa algoritma Naïve Bayes efektif dalam menganalisis opini publik pada isu sensitif di media sosial. Selain itu, hasil penelitian ini dapat menjadi acuan bagi pemerintah dan pemangku kebijakan dalam memahami persepsi publik serta merumuskan strategi komunikasi yang lebih tepat terkait kebijakan pendidikan.

Kata kunci: Sentimen Analisis, Naïve Bayes, Menteri Keuangan, Instagram, Analisis Teks.

PENDAHULUAN

Pendidikan merupakan salah satu pilar utama dalam pembangunan suatu

bangsa. Peran guru dan dosen tidak hanya terbatas pada proses transfer ilmu pengetahuan, tetapi juga mencakup pembentukan karakter, pengembangan potensi, serta persiapan generasi penerus yang berkualitas. Oleh karena itu, keberadaan tenaga pendidik memiliki posisi strategis dalam peningkatan kualitas sumber daya manusia. Namun demikian, isu-isu terkait apresiasi, penghargaan, dan dukungan terhadap profesi ini kerap menjadi perdebatan di ruang publik. Salah satu polemik yang sempat menjadi sorotan nasional adalah pernyataan Menteri Keuangan Republik Indonesia mengenai gaji Guru dan Dosen.

Pernyataan tersebut memicu beragam reaksi masyarakat, mulai dari dukungan, kritik, hingga kemarahan. Publik menilai bahwa ucapan tersebut dapat memengaruhi citra tenaga pendidik dan berpotensi menurunkan motivasi serta kepercayaan diri mereka. Reaksi masyarakat semakin meluas melalui media sosial, yang memungkinkan setiap individu menyuarakan opininya secara cepat, terbuka, dan masif. Fenomena ini menjadi contoh nyata bagaimana sebuah pernyataan pejabat publik mampu menimbulkan gelombang respons yang beragam dan sulit diprediksi.

Media sosial, khususnya Instagram, menjadi salah satu ruang diskusi utama dalam menanggapi isu tersebut. Indonesia merupakan salah satu negara dengan pengguna Instagram terbesar di dunia, dengan jumlah pengguna aktif mencapai lebih dari 100 juta pada tahun 2024. Karakteristik Instagram yang berbasis komentar singkat, cepat, dan real-time menjadikannya wadah yang efektif untuk menyampaikan pendapat. Namun, sifatnya yang terbuka juga memungkinkan munculnya opini yang ekstrem, baik positif maupun negatif, bahkan tanpa dukungan fakta yang memadai. Hal ini membuat analisis sentimen menjadi penting untuk memahami persepsi publik secara kuantitatif.

Analisis sentimen merupakan metode untuk mengidentifikasi dan

mengklasifikasikan opini publik ke dalam kategori tertentu, seperti positif, negatif, atau netral. Dengan analisis ini, peneliti dapat mengetahui kecenderungan emosi dan sikap masyarakat terhadap suatu topik atau pernyataan. Berbagai algoritma telah digunakan dalam analisis sentimen, salah satunya adalah Naïve Bayes. Algoritma ini populer karena sederhana, efisien, dan terbukti efektif dalam klasifikasi teks dengan jumlah data besar.

Algoritma Naïve Bayes bekerja berdasarkan teorema Bayes dengan asumsi independensi antar fitur. Meskipun asumsi ini sering kali tidak sepenuhnya terpenuhi dalam data teks, Naïve Bayes terbukti efektif dalam berbagai penelitian analisis sentimen. Keunggulannya terletak pada kemampuannya memproses data dengan cepat dan tetap menghasilkan akurasi yang kompetitif dibandingkan metode lain. Oleh karena itu, penggunaan algoritma ini relevan untuk mengolah data opini publik yang bersumber dari media sosial dengan jumlah yang masif.

Penelitian mengenai analisis sentimen di Indonesia sebelumnya banyak difokuskan pada isu politik, pelayanan publik, atau kebijakan pemerintah secara umum. Namun, kajian mengenai reaksi publik terhadap pernyataan pejabat negara yang menyinggung profesi strategis seperti guru dan dosen masih jarang dilakukan. Padahal, isu ini memiliki implikasi yang signifikan bagi dunia pendidikan dan kebijakan negara. Analisis sentimen dapat memberikan gambaran objektif mengenai persepsi publik, yang dapat dijadikan masukan dalam pengambilan keputusan dan strategi komunikasi pemerintah.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis sentimen pengguna Instagram terhadap pernyataan Menteri Keuangan tentang kebijakan gaji guru dan dosen menggunakan algoritma Naïve Bayes. Penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan metode analisis opini publik di media sosial, khususnya dalam

konteks isu pendidikan. Selain itu, hasil penelitian dapat menjadi masukan bagi pemangku kebijakan untuk memahami respons masyarakat serta merancang strategi komunikasi publik yang lebih efektif dan sensitif terhadap isu-isu yang melibatkan profesi pendidik.

Dengan demikian, penelitian ini tidak hanya memberikan nilai akademis dalam pengembangan analisis sentimen berbasis machine learning, tetapi juga memiliki relevansi praktis dalam mendukung pengambilan kebijakan publik. Di tengah era digital yang serba cepat, memahami opini publik secara akurat menjadi salah satu kunci untuk menjaga kepercayaan masyarakat terhadap pemerintah dan institusi negara.

METODE

Penelitian ini dilakukan langkah demi langkah, pada setiap alurnya dilakukan menggunakan metode yang kritis dan mampu mengarahkan pada proses yang sesuai sasaran. Metode penelitian ini merupakan cara-cara yang dipakai dalam merumuskan, menganalisa, mengumpulkan data sampai dengan tahap implementasi hasil penelitian. Metode penelitian disajikan pada gambar 1. Penelitian ini menggunakan dataset yang berasal dari Instagram dengan query yang berkaitan dengan pernyataan Menteri keuangan tentang kebijakan gaji guru dan dosen pada postingan Instagram dengan nama akun “nrasinewsroom”, dan menghasilkan data sebanyak 939 data, yang diuraikan menjadi 80% data training dan 20% data uji.

Pengumpulan data

Pada tahap awal penelitian ini diawali dari proses collecting data. Data diperoleh melalui proses scrapping pada media sosial dengan menggunakan Instant Data Scraping yang disediakan oleh google chrome dan sudah terinstal. Setelah mendapatkan data komentar maka selanjutnya mencari data yang relevan dengan penelitian.

Remove Duplicate

Pada tahap kedua ini peneliti melakukan pengecekan atau eliminasi data double dari hasil proses scrapping. Selanjutnya data tersebut akan dijadikan dataset yang akan dipakai untuk penelitian.

Preprocessing

Data set yang sudah ada perlu dilakukan processing untuk mengurangi noise dan pembersihan data sebelum dilakukan proses pengolahan data lebih lanjut. Dalam tahap ini ada beberapa tahapan yang harus dilakukannya diantaranya :

1. Mengubah Kata menjadi Lower adalah tahapan dimana semua huruf besar menjadi huruf kecil. Hal ini berfungsi menyamakan bentuk kata, mengurangi redundansi, meningkatkan akurasi model naïve bayes serta mempermudah tokenisasi dan vector.
2. Normalisasi, pada tahap ini dimana karakter dan tanda baca yang tidak berpengaruh dihapus dari teks. Bertujuan untuk mengurangi noise pada data set.
3. Stopword, adalah proses penghapusan pada se kalimat jika ada kata / term yang tidak dianggap terlalu penting. Oleh karena itu tujuan stopword removal ini untuk menghilangkan / menghapus kata / term yang dianggap tidak penting seperti waktu, penghubung, dan lain sebagainya agar memudahkan dalam proses klasifikasi
4. Tokenisasi adalah proses pemotongan string input berdasarkan tiap kata yang menyusunnya. Data komentar yang diambil akan dipotong menjadi sebuah token yang dipisahkan dengan tanda petik dan koma.

Visualisasi kata

Pada tahap ini dimana peneliti dapat menampilkan kata yang paling sering muncul, menampilkan kata dengan ukuran berbeda sesuai frekuensinya,

menampilkan kata-kata penting berdasarkan bobot TF-IDF. Hal ini bertujuan mengetahui kata apa saja yang sering muncul pada setiap kelas sentiment.

Preparation Data

Pada tahap ini peneliti melakukan pemisahan data menjadi data latih dan data uji, mengubah teks menjadi fitur numerik, melakukan oversampling menggunakan smote pada data latih. Hal ini bertujuan agar data lebih bersih dan konsisten dan mengurangi noise yang siap digunakan proses analisis.

Testing

Data latih yang telah melalui beberapa tahap akan menjadi input pada proses latih dalam implementasi ke dalam algoritma naïve bayes. Setelah melalui tahap perhitungan frekuensi pada tiap-tiap dokumen, maka dilakukan tahap pengklasifikasian memakai algoritma naïve bayes.

Kesimpulan

Memperoleh hasil dari perhitungan dengan algoritma naïve bayes dan memberikan kesimpulan dari penelitian ini.

HASIL DAN PEMBAHASAN

Pada tahap ini peneliti mengumpulkan data menggunakan metode scrapping dengan menggunakan ekstensi instant data scraper pada Google Chrome, data-data komentar tersebut diambil dari postingan Instagram akun narasinewroom mengenai gaji guru dan dosen, dari hasil proses *scrapping* data tersebut didapatkan sejumlah 939 data. Data tersebut kemudian disimpan dalam format .csv untuk dilakukan tahapan selanjutnya. Data hasil scrapping ini masih dalam bentuk yang tidak terstruktur dan mengandung banyak noise, seperti duplikasi data, tanda baca, angka, simbol, dan kata kata tidak baku yang tidak penting dalam tahapan klasifikasi. Maka

dari itu, data tersebut harus melalui remove duplicates dan tahap Pre-Processing untuk pengolahan lebih lanjut.

Tabel 1 Hasil Cleansing Data

no	nama_akun	text_cleaning	sentimen
0	_ayuhkoenyil	Memang gak punya malu , gitu aja ☐	Negatif
1	superbeng1988	Partisipasi dari masyarakat. itu penjelasan ...	Positif
2	haryosinggih	☐ mulai melawak☐	Positif
3	erdinp_17	☐Pala muyeng liat pemerintahan ☐	Positif

Remove Duplicates

Pada proses ini peneliti menghilangkan komentar yang mempunyai kesamaan kata agar kata bersifat unik, mnghindari penambahan bobot sampel serta mencegah sifat bias pada data.

Preprocessing

Tahapan ini dilakukan setelah collecting data adalah preprocessing. Pada tahap ini dilakukan pembersihan data yang tujuannya unruk mengurangi noise di dalam dataset. Sehingga proses minning akan berjalan maksimal.

Cleansing data

Pada tahap ini dilakukan proses menghilangkan semua data yang tidak dibutuhkan, misalnya karakter, tanda baca, tautan, tagar, mention dan angka, dan harapannya data yang akan digunakan dalam proses minning dalam keadaan bersih. Berikut codingan text cleansing data:

```
def clean_text(text):
    text = text.lower() # ubah ke huruf kecil
    text
```

```

=
re.sub(r"http\S+|www\S+|https\S+", "", text, flags=re.MULTILINE) # hapus URL
text = re.sub(r'@\w+|#\w+', "", text)

# hapus mention & hashtag
text = re.sub(r'[^a-zA-Z\s]', "", text)
# hapus angka, tanda baca, emoji
text = re.sub(r'(\d+|,|.|!|\?|\*|\(|\)|\{|\}|\[|\]|_|`|~|&|&#\d+|&#\w+)', r'\1', text)
# hilangkan huruf berulang (contoh: baguuuss -> bagus)
text = re.sub(r's+', ' ', text).strip()
# hapus spasi berlebih
return text
    
```

Tabel 2 Hasil Cleansing Data

No	Data Asli	Hasil
1	. Memang gak punya malu , gitu aja ðŸ˜	memang gak punya malu gitu aja
2	"Partisipasi dari masyarakat." itu penjelasan ...	partisipasi dari masyarakat itu jelas lengkap ...
3	ðŸ˜—¿ Mulai melawakðŸ˜—¿	mulai lawak

Mengubah data menjadi lower

Pada tahap ini dilakukan proses mengubah huruf kapital menjadi huruf kecil semua, hal ini bertujuan agar data yang akan digunakan dalam proses minning dalam keadaan seragam dan data menjadi efektif. Berikut codingan text proses merubah huruf kapital menjadi lower:

```

data['text_cleaning'] = data['text_cleaning'].str.lower()
data.head()
    
```

Tabel 3 Hasil Mengubah menjadi Lower

No	Komentar asli	Komentar setelah huruf menjadi lower
1	. Memang gak punya malu , gitu	. memang gak punya malu , gitu

	aja ☐	aja ☐
2	"Partisipasi dari masyarakat." itu penjelasan ...	"partisipasi dari masyarakat." itu penjelasan ...
3	☐ Mulai melawak☐	☐ mulai melawak☐

Tokenizing

Pada tahap ini dilakukan proses pemotongan string input berdasarkan tiap kata yang menyusunnya. Data komentar yang diambil akan dipotong menjadi sebuah token yang pisahkan dengan tanda petik dan koma. Berikut codingan text proses tokenizing:

```

# Tokenisasi
tokenized
data['text_cleaning'].apply(lambda x:x.split())
tokenized
    
```

Tabel 4 Hasil Tokenisasi

No	Komentar	Tokenisasi
1	. Memang gak punya malu , gitu aja ☐	[., memang, gak, punya, malu, ,, gitu, aja, ☐]
2	"Partisipasi dari masyarakat." itu penjelasan ...	["partisipasi, masyarakat.", penjelasan, lengk...
3	☐ Mulai melawak☐	[☐, mulai, melawak☐]

Stopword

Pada tahap ini dilakukan pemindaian semua data komentar dari awal sampai akhir dan kalua ada kata / term yang tidak dianggap tidak terlalu penting akan dihilangkan contohnya kata penghubung, waktu dan lain sebagainya tuhuannya untuk memudahkan dalam proses klasifikasi. Berikut codingan text proses stopwords:

```

# Stopword
import Sastrawi
    
```

```

from
Sastrawi.StopWordRemover.StopWordRe
moverFactory import
StopWordRemoverFactory,
StopWordRemover, ArrayDictionary
    more_stop_word = []stop_words =
StopWordRemoverFactory().get_stop_wo
rds()
new_array =
ArrayDictionary(stop_words)
    stop_words_remover_new =
StopWordRemover(new_array)

def stopword(str_text):
    str_text =
stop_words_remover_new.remove(str_tex
t)
    return str_text

data['text_cleaning'] =
data['text_cleaning'].apply(lambda x:
stop_words_remover_new.remove(x))
data.head()
    
```

Tabel 5 Hasil Stopword

No	Komentar	Stopword
1	"partisipasi dari masyarakat." itu penjelasan	"partisipasi masyarakat." penjelasan lengkapny...
	...	

Stemming

Pada tahap ini peneliti melakukan pemindaian semua data komentar dengan menghilangkan kata yang berimbuhan pada masing-masing kata sehingga menjadi kata dasar. Tujuannya untuk memudahkan proses klasifikasi. Berikut codingan text proses stemming:

```

# Terapkan stemming ke kolom
"Komentar"
data["Komentar_Stem"] =
data["text_cleaning"].astype(str).apply(ste
mming)
# Simpan hasil ke file baru
data.to_csv("Data_Komentar.csv",
index=False, encoding="latin1")
# Cek hasil
print(data.head(100))
    
```

Tabel 6 Hasil Stemming

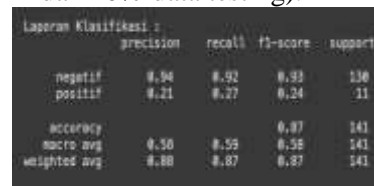
No	Komentar	Stemming
1	"Partisipasi dari masyarakat." itu penjelasan	partisipasi dari masyarakat itu jelas lengkap

Implementasi Naïve Bayes

Pada tahap ini peneliti melakukan penerapan naïve bayes Multinomial Naïve Bayes yang merupakan salah satu algoritma Naïve Bayes yang di implementasikan menggunakan library sklearn pada Python untuk dilakukan pengklasifikasian. Algoritma Naïve Bayes merupakan algoritma untuk metode klasifikasi dengan menjadi data set menjadi data training dan data testing.

Pengujian Klasifikasi

Pada penelitian ini, data diproses menggunakan algoritma Naïve Bayes dengan pembagian data 80:20 (80% data trainini dan 20% data testing).



Gambar 1 Laporan Klasifikasi Visualisassi Kata

Visualisasi Kata Negatif

Pada tahap ini peneliti melakukan visualisasi kata-kata negatif yang ada di komentar dalam bentuk grafik atau gambar agar lebih mudah dipahami. Berikut Proses dan Hasil Visualisasi Kata Negatif:

```

all_text_sO = ' '.join(word for word
in data_negatif['text_cleaning'])
wordcloud =
WordCloud(colormap='Reds',
width=1000, height=1000,
mode='RGBA',
background_color='white').generate(all_te
xt_sO)
plt.figure(figsize=(9, 6))
plt.imshow(wordcloud,
interpolation='bilinear')
plt.axis('off')
plt.title("Visualisasi Kata Negatif")
plt.margins(x=0, y=0)
    
```


- W. Yulita, E. D. Nugroho and M. H. Algifari. (2021) "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier," *JDMSI*, Vols. 2, No. 2, <https://ejurnal.teknokrat.ac.id/index.php/JDMSI/article/view/1344/672>, pp. 1-9, 2021.
- A. Safira and F. N. Hasan. (2023) "ANALISIS SENTIMEN MASYARAKAT TERHADAP PAYLATER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER," *Jurnal Sistem Informasi*, Vols. 5, No.1, <https://journal.unilak.ac.id/index.php/zn/article/view/12856>, pp. 59-70.
- S. Puad, G. and A. S. Y. Irawan. (2023) "ANALISIS SENTIMEN MASYARAKAT PADA TWITTER TERHADAP PEMILIHAN UMUM 2024 MENGGUNAKAN ALGORITMA NAÏVE BAYES," *Jurnal Mahasiswa Teknik Informatika*, Vols. 7, No.3, <https://ejournal.itn.ac.id/index.php/jati/article/view/6920>, Juni 2023.
- M. W. A. Putra, S. E. and H. (2020) "Analisis Sentimen Dompot Elektronik Pada Media Sosial Twitter Menggunakan Naïve Bayes Classifier," *IT Journal Research and Development (ITJRD)*, Vols. 5, No.1, 10.25299/itjrd.2020.vol5(1).5159, p. 72 – 86, Agustus 2020.
- R. W. Utami, A. Jazuli and T. Khotimah. (2021) "ANALISIS SENTIMEN TERHADAP XIAOMI INDONESIA MENGGUNAKAN METODE NAÏVE BAYES," *Indonesian Journal of Technology, Informatics and Science (IJTIS)*, Vols. 3, No. 1, 10.24176/ijtis.v3i1.7514, pp. 21-30, Desember 2021.
- W. Yulita. (2021) "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes," *Jurnal Data Mining dan Sistem Informasi*, vol. 2(2), pp. 1-9.
- F. Fatayat and R. Nugroho. (2021) "Analisa Penentuan Dosen Pembimbing Tugas Akhir Mahasiswa Menggunakan Naive Bayes Classifier," *Simtika*, vol. 4(3), pp. 1-7.
- A. Baita, Y. Pristyanto and N. Cahyono. (2021) "Analisis Sentimen Mengenai Vaksin Sinovac Menggunakan Algoritma Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN)," *Information System Journal*, vol. 4(2), pp. 42-46.
- K. D. Indarwati and H. Februariyanti. (2023) "ANALISIS SENTIMEN TERHADAP KUALITAS PELAYANAN APLIKASI GO-JEK MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER," *JATISI*, vol. 1, no. <https://doi.org/10.35957/jatisi.v10i1.2643>, p. 1.
- F. Fitriyadi dan Arsaela Astikasari. (2024) "Analisis Sentimen Masyarakat Terhadap Kebijakan Kenaikan Umk 6,5% Menggunakan Metode Naive Bayes" *RESTIA*, Vol. 3 No. 1, Hal 26-35.
- Kompas. (2023). Pernyataan Menteri Keuangan soal gaji guru dan dosen menuai pro-kontra. Kompas.com. <https://www.kompas.com>
- Nasrullah, R. (2017). Media sosial: Perspektif komunikasi, budaya, dan sosioteknologi. Simbiosis Rekatama Media.