

Analisis Sentimen Pada Kenaikan Gaji PNS, Guru, dan TNI -Polri Menggunakan Metode Naïve Bayes

Ichsan Azala¹, Yoga Firnanda²

^{1,2}Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

E-mail: ¹Ichsanmartin@gmail.com, ²firmanday89@gmail.com.

Abstrak – Kebijakan kenaikan gaji bagi guru, Pegawai Negeri Sipil (PNS), serta TNI–Polri merupakan isu strategis yang berdampak pada kualitas pendidikan, pelayanan publik, serta stabilitas pertahanan dan keamanan nasional. Kebijakan ini memunculkan beragam respons dari masyarakat yang banyak disampaikan melalui media sosial, khususnya platform YouTube. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kebijakan kenaikan gaji guru, PNS, dan TNI–Polri menggunakan metode Naïve Bayes. Data penelitian berupa komentar pengguna YouTube yang diperoleh melalui teknik web scraping menggunakan bahasa pemrograman Python. Tahapan penelitian meliputi pengumpulan data, preprocessing teks (case folding, tokenizing, stopword removal, filtering, dan stemming), pelabelan data, pembobotan kata menggunakan metode TF-IDF, proses klasifikasi menggunakan algoritma Naïve Bayes, serta evaluasi model menggunakan confusion matrix. Hasil pengujian dengan pembagian data training sebesar 80% dan data testing sebesar 20% menunjukkan bahwa metode Naïve Bayes menghasilkan nilai akurasi sebesar 83,4%, precision sebesar 78%, recall sebesar 83%, dan f1-score sebesar 79%. Hasil tersebut menunjukkan bahwa metode Naïve Bayes cukup efektif dalam mengklasifikasikan sentimen masyarakat terhadap kebijakan kenaikan gaji aparatur negara berdasarkan komentar di platform YouTube.

Kata Kunci — Analisis Sentimen, Naïve Bayes, Kenaikan Gaji, YouTube, TF-IDF

1. PENDAHULUAN

Kenaikan gaji bagi guru, PNS, dan TNI-Polri merupakan kebijakan penting yang berdampak pada kualitas pendidikan, pelayanan publik, serta pertahanan dan keamanan nasional. Aparatur negara memiliki peran strategis dalam pembangunan, sehingga perubahan kebijakan terkait kesejahteraan mereka selalu menjadi perhatian publik. Namun, respons masyarakat terhadap kebijakan tersebut sangat beragam, dipengaruhi oleh kondisi sosial-ekonomi dan persepsi mengenai prioritas anggaran negara.

Dengan berkembangnya teknologi digital, media sosial menjadi ruang utama bagi masyarakat untuk menyampaikan opini terkait kebijakan pemerintah. Beragam komentar yang muncul memperlihatkan adanya persepsi positif, negatif, maupun netral. Oleh karena itu, analisis sentimen menjadi pendekatan yang relevan untuk memahami bagaimana kebijakan kenaikan gaji aparatur negara diterima oleh publik. Salah satu metode klasifikasi yang banyak digunakan dalam analisis sentimen adalah *Naïve Bayes*, karena kemampuannya dalam menangani teks berdimensi tinggi serta menghasilkan akurasi yang kompetitif.

Penelitian terdahulu menunjukkan bahwa analisis sentimen dapat memberikan gambaran yang jelas mengenai kecenderungan opini masyarakat. Studi oleh [1] serta [2] menemukan bahwa sentimen publik terhadap kenaikan gaji guru cenderung positif. Sementara penelitian mengenai persepsi terhadap profesi PNS dan isu TNI, seperti yang dilakukan [3], menunjukkan pola sentimen yang lebih beragam dan sering kali kritis. Metode *Naïve Bayes* sendiri telah digunakan secara luas pada penelitian-penelitian tersebut karena kesederhanaan dan efektivitasnya dalam memproses data media sosial.

Beberapa penelitian sebelumnya menunjukkan bahwa metode *Naïve Bayes* cukup andal dalam menganalisis opini publik terkait kebijakan pemerintah. Studi oleh [4] menemukan bahwa mayoritas sentimen terkait kenaikan gaji guru honorer bersifat positif, dengan model Multinomial Naïve Bayes mencapai akurasi sekitar 89,65%. [5] juga menunjukkan bahwa *Naïve Bayes* efektif dalam mengidentifikasi sentimen masyarakat terhadap kebijakan kenaikan upah minimum daerah. Selain itu, studi komparatif oleh [6] membuktikan bahwa *Naïve Bayes* memiliki kinerja lebih baik dibandingkan SVM dan *Random Forest* dalam analisis sentimen kebijakan perumahan TAPERA, dengan akurasi sekitar 69,17%. Temuan serupa ditunjukkan oleh Setiawan [7] yang melaporkan akurasi hingga 88,89% pada analisis opini publik terhadap tokoh politik menggunakan dataset tidak seimbang.

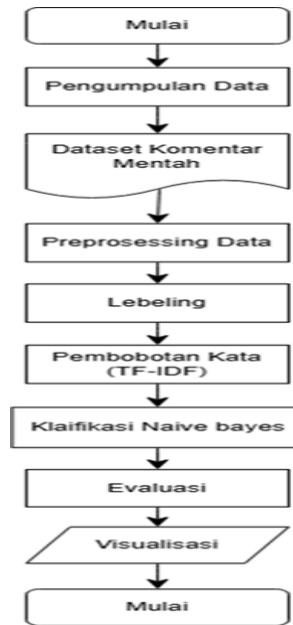
Meskipun demikian, sebagian besar penelitian sebelumnya hanya berfokus pada satu kelompok aparatur atau satu platform media sosial, sehingga belum memberikan gambaran komparatif yang menyeluruh. Berdasarkan hal tersebut, penelitian ini dilakukan untuk menganalisis dan membandingkan sentimen masyarakat terhadap kebijakan kenaikan gaji guru, PNS, dan TNI-Polri menggunakan metode *Naïve Bayes*. Hasil penelitian diharapkan

dapat memberikan masukan yang lebih objektif bagi pemerintah dalam merumuskan kebijakan kesejahteraan aparatur negara.

2. METODE PENELITIAN.

2.1 Diagram alur penelitian

Terdapat beberapa tahapan dalam penelitian kali ini diantaranya tahap pengumpulan data, preprocessing data, labeling, pembobotan kata, pemodelan metode klasifikasi, dan terakhir tahap evaluasi, sebagaimana diilustrasikan pada Gambar 1



Gambar 1 diagram alur penelitian

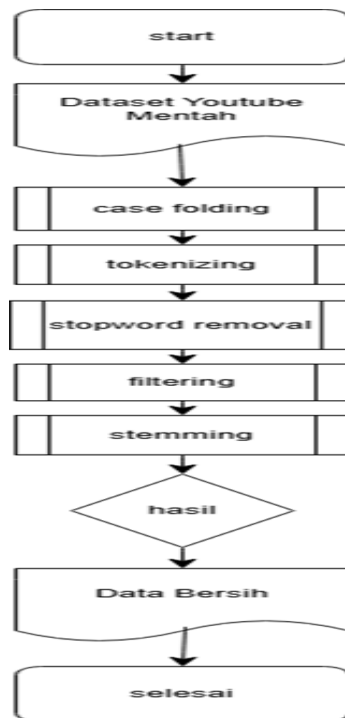
Penelitian berupa kumpulan komentar atau data yang didapat dari komentar *youtube* pada beberapa video yang membahas tentang “Kenaikan gaji Pns, Guru, Dan Tni -Polri”. Dalam melakukan penelitian ini diperlukan alur proses penelitian, agar dalam proses penelitian berjalan sesuai dengan rencana.

2.2 Pengumpulan Data

Tahap pengumpulan data dengan melakukan *scrapping* melalui komentar *youtube* dengan menggunakan modul *selenium* pada Bahasa pemrograman *Python*. Data yang di *scrap* adalah data yang bersumber dari video pembahasan “Kenaikan gaji Pns, Guru, Dan Tni -Polri”. Setelah itu dataset yang masi mentah akan disimpan dalam bentuk CSV (*Comma Separated Values*) yang nantinya akan diproses lebih lanjut.

2.3 Text Preprocessing

Proses pengumpulan data dikarenakan dataset tersebut termasuk dalam *unstructured* data (data tidak terstruktur). Sebelum dilakukan analisis lebih lanjut, terlebih dahulu dataset melalui proses *text preprocessing* untuk menghilangkan serta mengatasi noisy data agar hasil perhitungan optimal. Adapun alur proses *text preprocessing* antara lain, *case folding*, *tokenizing*, *stopword removal*, *filtering*, dan *stemming*. setelah data bersih maka dilakukan proses pelabelan secara manual. Berikut adalah alur diagram dari *text preprocessing*.



Gambar 2 diagram alur *preprocessing* data

- a. *Case Folding*
Case folding merupakan proses yang dilakukan untuk mengubah setiap kata yang ada di dalam dataset menjadi huruf kecil menggunakan fungsi *lowercase*.
- b. *Tokenizing*
Tokenizing adalah proses pemecahan teks menjadi unit-unit yang lebih kecil yang disebut token. Token dapat berupa kata, frasa, atau bahkan karakter, tergantung kebutuhan analisis. Pada tahap ini, kalimat dalam dataset dipecah menjadi daftar kata sehingga dapat diproses lebih lanjut oleh sistem.
- c. *Stopword Removal*
Proses *stopword removal* ini digunakan untuk menghilangkan kata yang tidak berpengaruh dalam proses sentimen. Kata dari data komentar you tube dibandingkan dengan kata yang terdapat dalam database *stopword*, hasil dari proses ini adalah menghilangkan kata yang terdeteksi sama pada kata *stopword* yang ada pada database.
- d. *Filtering*
Filtering merupakan tahapan yang bertujuan untuk menghilangkan kata umum yang biasa sering muncul dalam jumlah banyak dan tidak memiliki makna menggunakan algoritma *stoplist*(membuang kata yang kurang penting) atau *wordlist*(menyimpan kata penting).
- e. *Stemming*
Stemming adalah proses mengubah kata berimbuhan menjadi bentuk dasar (*root word*) dengan menghilangkan awalan, akhiran, atau sisipan yang ada pada kata tersebut. Tujuan *stemming* adalah menyamakan berbagai variasi kata yang memiliki makna serupa agar diproses sebagai satu kata dasar.

2.4 Pembobotan tf-idf

Pada pembobotan kata dilakukan proses pemberian nilai atau bobot terhadap setiap kata pada komentar you tube yang telah melewati tahap *preprocessing*. Metode yang digunakan dalam pemberian bobot pada term yaitu TF-IDF. Pembobotan ini bertujuan untuk memberikan nilai kepada sebuah *term* dimana nilai dari *term* tersebut akan dijadikan sebagai input pada proses klasifikasi.

$$TF - IDF = TD * IDF..... (1)$$

Dimana:

$$TF \frac{\text{jumlah kata pada dokumen}}{\text{total kata pada dokumen}} (2)$$

$$IDF \frac{\text{total dokumen}}{\text{jumlah dokumen yang mengandung kata}} (3)$$

Keterangan:

- TF: Menggambarkan seberapa sering sebuah *term* muncul dalam satu dokumen tertentu.
- IDF: berfungsi untuk menurunkan bobot kata-kata yang terlalu sering muncul di banyak dokumen dan menaikkan bobot kata yang jarang muncul.
- TD (TF-IDF): merupakan hasil perkalian antara nilai TF dan IDF.

2.5 Klasifikasi Naive Bayes

Setelah dilakukan proses pembobotan TF IDF, selanjutnya masuk ke tahap analisis data menggunakan algoritma yang telah ditentukan yaitu *Naive Bayes*. Pada tahap ini data dibagi menjadi dua bagian, yaitu testing dan training. Dalam tahap ini juga akan dilakukan proses klasifikasi untuk menentukan label positif dan negatif dari *term* yang sudah melewati proses pembobotan TF IDF untuk mendapatkan nilai akurasi dari algoritma *Naive Bayes*.

2.6 Evaluasi

Evaluasi dalam penelitian ini bertujuan untuk mengetahui kinerja atau performa dari model yang diusulkan. Metode yang digunakan untuk evaluasi pada penelitian ini adalah *confusion matrix*. *Confusion matrix* merupakan sebuah tabel yang memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem (prediksi) dengan hasil klasifikasi yang sebenarnya. Dalam *confusion matrix*, terdapat 4 istilah nilai hasil klasifikasi yang nantinya akan digunakan untuk menghitung nilai akurasi, presisi, *recall*, dan *f1-score*, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).

$$Accuracy \frac{(TP+FN)}{(TP+FP+FN+TN)} (4)$$

$$Precision \frac{(TP)}{(TP+FP)} (5)$$

$$Recall = \frac{(TP)}{(TP+FN)} (6)$$

3. HASIL DAN PEMBAHASAN

3.1 Preprocessing

1. Case Folding

Case folding dilakukan untuk mengubah seluruh ukuran huruf pada kata menjadi suatu bentuk ukuran huruf yang sama. Karena tidak semua tweet konsisten dalam penggunaan ukuran huruf. Fungsi *case folding* ini akan mengubah semua huruf menjadi *lowercase* atau huruf kecil. Hasil dari proses *case folding* disajikan pada table 1

Tabel 1 *case folding*

| komentar | Case folding |
|--|--|
| gak akan ngefek jk para korup tidak dimiskinkan ingat partai yg gak kerja u rakyat jgn pilih | gak akan ngefek jk para korup tidak dimiskinkan ingat partai yg gak kerja u rakyat jgn pilih |
| Menku nya bingung nya cari uang 🤔 | Menku nya bingung nya cari uang |

2. Tokenizing

Proses *tokenizing* dilakukan untuk memisahkan deretan kata di dalam kalimat, paragraf atau halaman menjadi token atau potongan kata tunggal. Pada saat bersamaan, tokenizing juga membuang karakter selain huruf seperti tanda baca. Hasil dari proses tokenization disajikan dalam Tabel 2

Tabel 2. Tokenizing

| komentar | Tokenizing |
|--|---|
| saya klo ingat kata kata mulyani itu saya jenenggakel bngt mulutnya ngomong ga sadar yang bikin kamu pinter itu siapa mulyani raimu itu koh bikin rakyat indsatusia marah gara gara ucapanmu mulyani | ['saya', 'klo', 'ingat', 'kata', 'kata', 'mulyani', 'itu', 'saya', 'jenenggakel', 'bngtmulutnya', 'ngomong', 'ga', 'sadar', 'yang', 'bikin', 'kamu', 'pinter', 'itu', 'siapa', 'mulyani', 'raimu', 'itu', 'koh', 'bikin', 'rakyat', 'indsatusia', 'marah', 'gara', 'gara', 'ucapanmu', 'mulyani'] |

3. Stopword removal

Proses *stopword removal* ini digunakan untuk menghilangkan kata yang tidak berpengaruh dalam proses sentimen. Kata dari data komentar you tube dibandingkan dengan kata yang terdapat dalam database *stopword*, hasil dari proses ini adalah menghilangkan kata yang terdeteksi sama pada kata *stopword* yang ada pada database. Hasil dari proses *stopword removal* dapat dilihat pada Tabel 3.

Tabel 3 *stopwrod removal*

| komentar | Stopword removal |
|--|--|
| bagus juga memang jika mau memajukan pendidikan begitu supaya guru mengajar semangat | ['bagus', 'memajukan', 'pendidikan', 'guru', 'mengajar', 'semangat'] |
| Presiden terburuk sepanjang sejarah plagi diskriminasi Bolot dasar guru jaman SBY dah besar klw sertipikasi 7 juta jadi 9 juta , hakim yang maling dikasih 280 persen macan ompong Bolot dasar | ['presiden', 'terburuk', 'sejarah', 'plagi', 'diskriminasi', 'bolot', 'dasar', 'guru', 'jaman', 'sby', 'dah', 'klw', 'sertipikasi', 'juta', 'juta', 'hakim', 'maling', 'dikasih', 'persen', 'macan', 'ompong', 'bolot', 'dasar'] |

4. Steaming

Pembobotan ini bertujuan untuk memberikan nilai kepada sebuah term dimana nilai dari *term* tersebut akan dijadikan sebagai *input* pada proses klasifikasi. Hasil dari proses pembobotan menggunakan TF-IDF didapatkan nilai sentimen pada setiap kalimat, dan disajikan pada Tabel 4

Tabel 4 *stemming*

| komentar | <i>Steaming</i> |
|--|--|
| Habis manis pensiunan dilupakan hidup omonomon | [habis manis pensiun lupa hidup omonomon] |
| ngomong aja kenyataannya bohooooong besar sampai sekarang juga mana omon omon kosong | ngomong aja kenyataannya bohooooong besar sampai sekarang juga mana omon omon kosong |

3.2 Pembobotan TF-IDF

Pembobotan ini bertujuan untuk memberikan nilai kepada sebuah *term* dimana nilai dari *term* tersebut akan dijadikan sebagai *input* pada proses klasifikasi. Hasil dari proses pembobotan menggunakan TF-IDF didapatkan nilai sentimen pada setiap kalimat

3.3 Klasifikasi dan pengujian

Klasifikasi menggunakan metode *Naive bayes* dilakukan dengan membandingkan setiap kata dalam teks dengan daftar kata berlabel positif dan negatif. Jika teks mengandung lebih banyak kata bernada positif, maka kalimat dikategorikan sebagai sentimen positif dan sebaliknya. Metode ini bekerja berdasarkan keberadaan dan frekuensi kata dalam kamus sentimen, sehingga penentuan kategori dilakukan secara langsung berdasarkan makna kata yang muncul dalam teks

```
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

model = MultinomialNB()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
```

Gambar 3

3.4 Evaluasi

Pengujian kinerja dari *Naive Bayes* dilakukan dengan metode *confusion matrix* berdasarkan data *testing* yang akan di proses oleh model Klasifikasi *Naive Bayes* yang telah dibuat sebelumnya. Metode *confusion matrix* menghasilkan empat nilai yaitu *True Positive*, *True Negative*, *False Positive* dan *False Negative*. Selanjutnya dilakukan perhitungan mulai dari *accuracy*, *precision* dan *recall*. Berikut adalah hasil dari pengujian dengan data *training* 80% dan data *testing* 20%, Hasil dari evaluasi algoritma *Naive bayes* disajikan pada tabel berikut

Tabel 5 hasil evaluasi

| Evaluasi | hasil |
|-----------|-------|
| Accuracy | 0.834 |
| Precision | 0.78 |
| Recall | 0.83 |
| F1 score | 0.79 |

4. SIMPULAN

Berdasarkan hasil dan pembahasan dapat ditarik kesimpulan mengenai analisis sentimen terhadap kenaikan gaji guru,pns, dan tni-polri di *Platform* You Tube menggunakan metode *Naïve Bayes* sebagai berikut:

- a. Penerapan metode *Naïve Bayes* untuk analisis sentimen terhadap kenaikan Gaji Guru,Pns, dan Tni-Polri. Peneliti memperoleh hasil evaluasi dengan menggunakan *Confusion Matrix*. Akurasi sebesar 83%, ketepatan antara hasil prediksi dengan data sebenarnya (precision) sebesar 78%, tingkat keberhasilan sistem dalam memprediksi sebuah data (recall) sebesar 83%, serta tingkat kesalahan semua data yang diprediksi (error rate) sebesar 17%. Nilai perbandingan rata-rata precision dan recall (f1-score) adalah sebesar 79%.
- b. Berdasarkan hasil dan pembahasan menunjukkan bahwa metode *Naïve Bayes* dapat diterapkan untuk analisis sentimen komentar di *Platform* You Tube.

5. SARAN

Penelitian selanjutnya, penelitian dapat mempertimbangkan penggunaan algoritma klasifikasi lain yang lebih sesuai dengan karakteristik pada dataset yang akan digunakan. Setiap metode memiliki karakteristik yang berbeda dalam menangani variasi teks, pola bahasa, serta distribusi kelas. Oleh karena itu, untuk penelitian mendatang, pertimbangkan untuk penggunaan metode lain yang dapat memberikan hasil yang lebih optimal.

DAFTAR PUSTAKA

- [1] M. Jamaluddin and E. I. Habibillah, "Pengaruh Kebijakan Perserikatan Bangsa-Bangsa (PBB) Dalam Konflik Palestina-Israel," *Jurnal Tapis: Teropong Aspirasi Politik Islam*, vol. 19, no. 1, pp. 1–23, 2020.
- [2] N. Hendrastuty, A. R. Isnain, and A. Y. Rahmadhani, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," *Jurnal Informatika*, 2021.
- [3] A. Surahman, "Pengembangan Market Segmentasi untuk Mencapai Keunggulan Bersaing pada E-Marketplace," *Jurnal Komputer dan Informatika*, vol. 15, no. 1, pp. 118–126, 2020.
- [4] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *Jurnal Media Informasi Budidarma*, vol. 4, no. 3, pp. 650–656, 2020, doi: 10.30865/mib.v4i3.2181.
- [5] A. Putra, D. Haeirudin, H. Khairunnisa, and R. Latifah, "Analisis Sentimen Masyarakat Terhadap Kebijakan PPKM Pada Media Sosial Twitter Menggunakan Algoritma SVM," *Seminar Nasional Sains dan Teknologi*, pp. 1–6, 2021.
- [6] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, "Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM)," *Jurnal Gaussian*, vol. 9, no. 3, pp. 376–390, 2020, doi: 10.14710/j.gauss.v9i3.28932.
- [7] F. D. Ananda and Y. Pristyanto, "Analisis Sentimen Pengguna Twitter Terhadap Layanan Internet Provider Menggunakan Algoritma Support Vector Machine," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 20, no. 2, pp. 407–416, 2021, doi: 10.30812/matrik.v20i2.1130.
- [8] M. I. Fikri, T. S. Sabrila, Y. Azhar, and U. M. Malang, "Comparison of the Naïve Bayes Method and Support Vector Machine on Twitter Sentiment Analysis," *SMATIKA: STIKI Informatics Journal*, vol. 10, no. 2, pp. 71–76, 2020.
- [9] M. A. Reinfeld, "A Case Study of the GRP 5000," *Technology*, vol. 5, no. 3123, pp. 548–554, 2011.
- [10] Rahmawati and Pratama, "Analisis Sentimen terhadap Kenaikan Gaji Guru Honorer Menggunakan Multinomial Naïve Bayes," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 1, 2023.
- [11] Suryani, D. Wahyu, and A. Lestari, "Analisis Sentimen Masyarakat Terhadap Kebijakan Kenaikan Upah Minimum Menggunakan Naïve Bayes," *Jurnal RESTIA*, 2022.
- [12] Hidayat and Kurniawan, "Comparison of Naïve Bayes, SVM, and Random Forest in Public Sentiment Analysis of TAPERA Policy," *International Journal of Science, Technology & Management*, vol. 6, no. 3, pp. 1220–1228, 2023.
- [13] Setiawan, "Analisis Sentimen Opini Publik Terhadap Tokoh Politik Menggunakan Naïve Bayes," *Yayasan Kita Menulis*, 2023.