

ANALISIS SENTIMEN PENILAIAN MASYARAKAT INDONESIA TERHADAP KONFERENSI G20 DI BALI DENGAN MENGGUNAKAN METODE *NAIVE BAYES*

Windy Kusumawati Suhet ^{1*}, Adi Nugroho ².

^{1*,2} Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Kota Salatiga, Provinsi Jawa Tengah, Indonesia.

Email: windykusumawatisuhet@gmail.com ^{1*}, adi.nugroho@uksw.edu ²

Histori Artikel:

Dikirim 23 Mei 2023; *Diterima dalam bentuk revisi* 15 Juni 2023; *Diterima* 1 Juli 2023; *Diterbitkan* 10 September 2023. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

Abstrak

Banyak dari masyarakat Indonesia mendukung diselenggarakannya konferensi G20 di Bali namun banyak juga masyarakat yang tidak setuju dengan adanya konferensi G20 yang diselenggarakan di Bali dengan berbagai alasan. Oleh karena itu penelitian ini dilakukan untuk mengetahui seberapa banyak masyarakat Indonesia setuju atau tidak setuju dengan diadakannya Konferensi Tingkat Tinggi (KTT) G20 yang diselenggarakan di Bali, Indonesia. Pada penelitian ini digunakan metode penelitian pengumpulan data, tahap preprocessing, tahap klasifikasi Naive Bayes, dan tahap evaluasi confusion matrix. Dalam penelitian ini digunakan 2.255 data dengan data latih sebanyak 1.000 data dan terdapat 1.255 data uji. Setelah data diolah dan telah diprediksi sentiment pada data uji dilakukan perhitungan confusion matrix untuk menghitung akurasi hasil. Pada data yang telah diproses accuracy yang diperoleh sebesar 95,60%.

Kata Kunci: Konferensi G20; Youtube; Naïve Bayes.

Abstract

Lot of Indonesian people support the holding of the G20 conference in Bali and many people also disagree with the G20 that will held/hold in Bali for various reasons. Therefore this research is conducted to find out how much the Indonesian people agree or disagree with the holding of the G20 High-Level Conference (Summit) which was held in Bali, Indonesia. In this study used research methods of data collection, preprocessing stage, Naive Bayes classification stage, and confusion matrix evaluation stage. In this study, 2,255 data were used with 1,000 training data and 1,255 test data. After the data is processed and the sentiment predicted in the test data, the confusion matrix is calculated to calculate the accuracy of the results. On data that has been processed the accuracy obtained is 95.60%.

Keyword: G20 Conference; Youtube; Naïve Bayes.

1. Pendahuluan

Perkembangan teknologi informasi saat ini tidak dapat dipisahkan lagi dari kehidupan sehari-hari. Teknologi informasi ini dapat melingkupi banyak bidang antara lain dalam hal: pendidikan, kesehatan, perbankan, bisnis maupun dalam mencari informasi, karena teknologi informasi dalam perkembangannya begitu cepat, secara tidak langsung mengharuskan manusia untuk menggunakan teknologi dalam segala hal, salah satunya adalah dalam mencari informasi. Dengan adanya teknologi orang-orang dapat secara cepat mendapatkan informasi yang terbaru baik dalam bentuk teks maupun video. Pada zaman ini banyak sekali media informasi yang dapat digunakan untuk mencari informasi salah satunya Youtube.

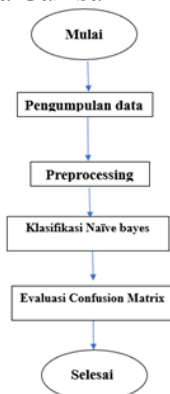
Youtube merupakan wadah bagi masyarakat luas untuk mencari informasi serta media *sharing* video online, selain itu kegunaan youtube digunakan sebagai media pembelajaran dan media informasi [1]. Pengguna Youtube dapat diakses dari berbagai kalangan, mulai dari orang dewasa hingga anak kecil. Banyak orang menggunakan Youtube untuk mengupload video, *search* video, menonton video, dan diskusi/tanya jawab mengenai video [2].

Pada tahun 2022 Indonesia menjadi tuan rumah Konferensi Tingkat Tinggi (KTT) G20 yang diselenggarakan di Bali pada tanggal 15 – 16 November 2022. Anggota G20 terdiri dari Afrika Selatan, Amerika Serikat, Arab Saudi, Argentina, Australia, Brasil, India, Indonesia, Inggris, Italia, Jepang, Jerman, Kanada, Meksiko, Republik Korea, Rusia, Perancis, Tiongkok, Turki, dan Uni Eropa [3]. Konferensi G20 merupakan forum internasional yang memiliki ekonomi maju di dunia yang berfokus dalam masalah penting perekonomian di dunia [4]. Konferensi G20 yang diadakan di Bali ini tidak hanya disajikan untuk tamu undangan tetapi ditayangkan juga melalui media sosial salah satunya youtube yang dapat dikonsumsi oleh masyarakat Indonesia.

Dalam tayangan youtube di channel KompasTv, konferensi yang sedang dilaksanakan di Indonesia berlangsung dengan rangkaian acara yang sangat megah dan mewah. Hal ini menyebabkan pro dan kontra terhadap masyarakat Indonesia, banyak dari masyarakat Indonesia mendukung dengan diselenggarakannya G20 di Bali dan banyak juga masyarakat yang tidak setuju dengan adanya G20 yang diselenggarakan di Bali dengan alasan rangkaian acara terlalu megah sehingga menguras keuangan negara. Oleh karena itu penelitian ini dilakukan untuk mengetahui opini masyarakat terhadap Konferensi Tingkat Tinggi G20 melalui komentar di youtube dan untuk mengetahui seberapa banyak masyarakat Indonesia setuju atau tidak setuju dengan diadakannya Konferensi Tingkat Tinggi (KTT) G20 yang diselenggarakan di Bali, Indonesia.

2. Metode Penelitian

Dalam penelitian ini akan dilakukan beberapa tahapan penelitian analisis sentiment penilaian masyarakat Indonesia terhadap G20 di Bali dengan menggunakan Metode Naive Bayes. Tahap penelitian yang akan dilakukan tertera pada Gambar 1.



Gambar 1. Tahapan penelitian

2.1 Pengumpulan Data

Tahapan pertama terdapat tahapan pengumpulan data, dimana pada tahapan ini akan dilakukan proses pengambilan data dari komentar youtube yang menayangkan cuplikan kegiatan G20, pengumpulan data akan menggunakan API youtube, hal ini dilakukan untuk mempermudah penelitian dalam pengumpulan data. API youtube nantinya akan mengambil data komentar dari server youtube [5]. Didalam proses pengumpulan data akan diadakan tahap pemfilteran data dengan menggunakan text mining. Tahapan text mining yang akan dilakukan yaitu melakukan input teks, kemudian teks yang sudah di input akan dilakukan pemfilteran sehingga akan terpilih kandidat teks yang akan digunakan. Ketika terpilih kandidat teks yang akan digunakan dilakukan pelabelan sentimen data terhadap komentar youtube. Untuk nilai label sentimen yang diberikan yaitu positif dan negatif [6].

2.2 Preprocessing Data

Selanjutnya dilakukan tahap *preprocessing* data, dimana akan dilakukan proses *cleansing* data [7]. *Processing* data dilakukan bertujuan untuk mempersiapkan data teks mentahan yang siap untuk di analisis teks sesuai dengan metode Naïve Bayes [8]. Dalam tahap preprocessing proses yang dilakukan yaitu *tokenize*, *transform cases*, *stopword filters*, dan *tokens by length* [9].

2.3 Klasifikasi Naïve Bayes

Selanjutnya adalah tahap klasifikasi Naïve Bayes, tahapan ini adalah proses pengklasifikasian data dengan bantuan aplikasi Rapidminer. Klasifikasi Naïve Bayes digunakan untuk memprediksi probabilitas atau suatu kelas keputusan [10]. Proses klasifikasi yang akan dilakukan dengan membuat model klasifikasi dan melakukan tahap prediksi sentimen terhadap data uji [11].

2.4 Evaluasi Confusion Matrix

Kemudian tahap evaluasi confusion matrix, pada tahap ini akan dilakukan evaluasi data dengan menggunakan rumus confusion matrix. Dalam tahapan confusion matrix digunakan untuk melakukan perhitungan akurasi atau tingkat kebenaran dari proses klasifikasi [12]. Proses evaluasi confusion matrix akan dilihat dari komentar masyarakat di video youtube KompasTv. Berikut rumus dan contoh tabel confusion matrix [13].

Tabel 1. Confusion Matrix

Correct Classification	Classification	
	Positif	Negatif
Positif	TP	TN
Negatif	FP	FN

$$\text{Accuracy} = \frac{TP+TN}{\text{Total}} \times 100 \%$$

$$\text{Precision} = \frac{TP}{TP+FP} \times 100 \%$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100 \%$$

Keterangan :

TP (*True Positif*) = jumlah data kelas positif dan terklasifikasi positif

TN (*True Negative*) = jumlah data kelas negatif dan terklasifikasi negatif

FN (*False Negative*) = jumlah data kelas negatif, tetapi terklasifikasi salah

FP (*False Positif*) = jumlah data kelas positif, tetapi terklasifikasi salah.

Tabel 2. Contoh data sebelum dihilangkan emoji

Name	Comment
Wirdawaty Hasibuan	sangat bagus, Saya sudah berkali-kali menontonnya tidak bosan-bosan, mantaap
Dolan Yok annarena	keren bgt Melihat senyum para pemimpin dunia "it's the beautiful night" ❤️❤️❤️MCMCMC👍👍👍
Nurul Widiastuti	Majulah negeriku Indonesia... Sebarkan kedamaian disetiap negri Bersama membangun bangsa
Risa Kurniawati	Indonesia the best love love ❤️❤️
Murni Murniati	Keren Indonesia 👍👍👍

Tabel 3. Contoh data yang telah dihilangkan karakternya

Name	Comment
Wirdawaty Hasibuan	sangat bagus, Saya sudah berkali-kali menontonnya tidak bosan-bosan, mantaap
Dolan Yok annarena	keren bgt Melihat senyum para pemimpin dunia "it's the beautiful night";
Nurul Widiastuti	Majulah negeriku Indonesia... Sebarkan kedamaian disetiap negri Bersama membangun bangsa
Risa Kurniawati	Indonesia the best love love
Murni Murniati	Keren Indonesia

Kemudian menghilangkan kata dan karakter yang tidak penting, salah satunya kata dan karakter
 yang sering muncul. Dalam hal ini digunakan operator *replace* pada Rapidminer. Setelah itu dilakukan proses *tokenize*, dimana hal ini digunakan untuk memisahkan atau memecahkan kata per-kata. Dalam prosesnya *tokenize* dilakukan didalam parameter operator *Process Document from Data* . Dalam parameter *tokenize*, *mode* yang dipilih yaitu *non letters*.

Selanjutnya dilakukan tahap *transform case*, hal ini dilakukan untuk menyamaratakan huruf besar menjadi huruf kecil. Pada tahap ini digunakan operator *transform case* pada Rapidminer, sebagai contoh pada gambar. terdapat kata Indonesia (huruf besar pada I nya) dan satu lagi dengan kata Indonesia huruf i kecil, sehingga Rapidminer membaca kedua kata merupakan data yang berbeda. Pada tahap *transform case* akan dipilih *lower case* sehingga semua huruf besar akan berganti menjadi huruf kecil.

Tabel 4. Contoh data yang belum di transform case

Word	Attribute Name	Total	Document
Indonesia	Indonesia	760	623
dan	dan	392	306
indonesia	indonesia	235	202

Tabel 5. Contoh data yang telah di transform case

Word	Attribute Name	Total	Document
indonesia	indonesia	1131	905
dan	dan	392	306

Kemudian, tahap filter *stopwords* yang digunakan untuk menyeleksi kata-kata yang tidak penting pada data. Pada rapidminer belum mendukung operator untuk berbahasa Indonesia, sehingga dalam hal ini digunakan operator filter *stopwords* (*Dictionary*). Saat melakukan filter *stopwords*, sebagai list data

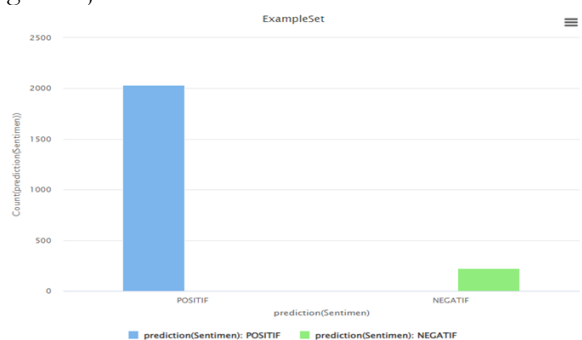
stopwords digunakan data yang diambil dari situs *website kaggle* sebagai data acuan. Dalam mengakses list *stopwords* di *website kaggle* dapat menggunakan *keyword Indonesia stoplist*. Setelah dilakukan *stopwords*, masih terdapat kata-kata yang mengandung banyak huruf dan mengandung kalimat yang disingkat, untuk permasalahan ini digunakan operator pada Rapidminer yaitu filter *tokens (by length)*. Dalam pengaturan filter ini akan digunakan minimal kata sebanyak 4 dan untuk maksimal kata sebanyak 25 kata, sehingga kata-kata yang tidak memenuhi syarat akan dihilangkan.

3.3 Klasifikasi

Klasifikasi digunakan untuk membangun suatu model data dari data latih yang akan di klasifikasi pada data baru [15], untuk penelitian ini terdapat 2.255 data dengan data latih sebanyak 1.000 (sudah diberi label) dan terdapat 1.255 data uji (yang belum diberi label). Dalam pemberian label ditelaah dan dilabeli secara manual. Tahap pertama akan dilakukan pembuatan model klasifikasi, dengan cara yaitu mengambil data latih untuk diproses terlebih dahulu. Dalam tahapan ini digunakan operator filter *examples* pada Rapidminer, kemudian diambil data sentiment dengan label yang tidak kosong (*is not missing*) atau data latih. Dalam pelabelan data terdapat dua sentiment yaitu positif dan negatif, dimana *comment* youtube akan ditelaah dengan sesuai dan diberi sentiment, yang dimana akan menjadi data latih, sedangkan data yang belum diberi sentiment dinamakan data uji.

Setelah itu masuk ke tahap *preprocessing* dimana dalam penelitian ini sudah dilakukan tahap *preprocessing* terlebih dahulu sehingga langsung dihubungkan operatornya. Kemudian, masuk ke tahap penggunaan algoritma Naïve Bayes yang akan digunakan untuk mempelajari pola data latih untuk menghasilkan model klasifikasi, sehingga dapat digunakan untuk menganalisa sentimen data yang belum diberi label. Setelah dilakukan proses pada algoritma Naïve Bayes dilakukan penyimpanan model Naïve Bayes untuk melakukan tahap selanjutnya. Model Naïve Bayes akan dipakai sebagai model untuk mengklasifikasi data uji yang belum berlabel berdasarkan data latih yang telah diproses. Pada tahap ini untuk menyimpan model yang telah dibuat digunakan operator *store*, selain menyimpan model Naïve Bayes dilakukan juga penyimpanan untuk data latih yang sudah diproses. Jika data sudah tersimpan, selanjutnya melakukan tahapan kedua yaitu untuk melabeli data uji yang belum ada labelnya. Tahapannya dengan mengambil data yang belum diberi label dengan cara yang sama menggunakan operator filter *example* dengan memilih data sentimen yang tidak memiliki label (*is missing*).

Setelah itu dilakukan tahapan seperti sebelumnya dengan menambahkan operator *nominal to text* dan *process document from data*. Kemudian, *drag and drop* data latih yang sebelumnya telah tersimpan. Supaya atribut data latih dan data uji tergabung menjadi satu maka digunakan operator *union*. Setelah atribut digabungkan dilakukan *replace missing value* dikarenakan ada beberapa atribut yang telah digabungkan memiliki data *missing*. Proses penggunaan operator *replace missing value* yaitu dengan mengubah default menjadi *zero* yang artinya nilai *missing* akan digantikan dengan nilai 0. Setelah dilakukan *replace missing value drag and drop* model klasifikasi yang telah dibuat dan disimpan kedalam lembar kerja. Setelah di *drag and drop* tambahkan pada lembar kerja operator *apply model*. Hasil dari *prediction(sentiment)* bisa dilihat dari visualisasi data, dalam tahap ini digunakan jenis visualisasi berupa bar(column). Dapat dilihat nilai *prediction(sentiment)* positif sejumlah 2.034 dan *prediction(sentiment)* negatif sejumlah 221.



Gambar 2. Hasil *prediction(sentimen)*

3.4 Evaluasi Konfusion Matrix

Setelah melakukan tahap klasifikasi dilakukan tahapan *confusion matrix* untuk menganalisa keakurasian hasil. Dalam tahap ini akan dilihat dari dua cara, yang pertama dapat dilihat dari operator performance pada Rapidminer, dapat dilihat pada gambar 7. bahwa akurasi yang diperoleh sebesar 95,60%.

accuracy: 95.60%

	true POSITIF	true NEGATIF	class precision
pred. POSITIF	934	0	100.00%
pred. NEGATIF	44	22	33.33%
class recall	95.50%	100.00%	

Gambar 3. Hasil *accuracy* data

Cara kedua dengan perhitungan manual untuk melihat nilai *recall*, *precision*, dan *accuracy* dengan acuan data pada hasil performance pada Rapidminer, sebagai berikut :

$$Recall = \frac{TP}{TP + FN} = \frac{934}{934 + 44} = 0,9550$$

$$Precision = \frac{TP}{FP + TP} = \frac{934}{0 + 934} = 1$$

$$Accuracy = \frac{TP + TN}{Total} = \frac{934 + 22}{1000} = 0.956$$

Dari hasil perhitungan, bahwa perhitungan *recall* untuk memprediksi komentar yang bersifat negative mendapat presentase 95,50 %, kemudian untuk perhitungan *precision* digunakan untuk mengidentifikasi nilai prediksi positif berdasarkan keseluruhan nilai prediksi negatif, sehingga didapat presentase sebesar 100%, dan untuk nilai *accuracy* yang didapatkan pada penelitian ini sebesar 95,60%.

4. Kesimpulan

Berdasarkan hasil penelitian Analisis Sentimen Penilaian Masyarakat Indonesia Terhadap Konferensi G20 di Bali dengan menggunakan Metode Naive Bayes, disimpulkan bahwa komentar masyarakat Indonesia di media sosial youtube cenderung positif. Dengan metode naïve bayes mampu mengklasifikasi sentimen dengan *accuracy* data dengan nilai presentase 95,60% berdasarkan 2.255 data yang terdiri dari 1.000 data latih dan 1.255 data uji.

5. Ucapan Terima Kasih

Terima kasih kepada teman-teman yang telah mendukung selama proses pembuatan jurnal ini dan terima kasih kepada kampus Universitas Kristen Satya Wacana terutama Fakultas Teknologi Informasi yang telah memfasilitasi dalam penyusunan artikel jurnal ini.

6. Daftar Pustaka

- [1] Mujiyanto, H. (2019). Pemanfaatan Youtube sebagai media ajar dalam meningkatkan minat dan motivasi belajar. *Jurnal Komunikasi Universitas Garut: Hasil Pemikiran dan Penelitian*, 5(1), 135-159.
- [2] Husna, A., Murdaningtyas, A. A., & Hajron, K. H. (2022). Meningkatkan Prestasi Belajar Siswa Menggunakan Audio Visual. *Prosiding Konferensi Ilmiah Dasar*, 3, 1327-1337.
- [3] Yanthi, N. P. D. M., Yudhaningsih, N. M., & Pering, I. M. A. A. (2022). Peluang Dan Tantangan SDM Bali Serta Strategi Dalam Menyambut Kegiatan Presidensi G20. *JURNAL EKONOMI, MANAJEMEN, BISNIS, DAN SOSIAL (EMBISS)*, 2(4), 633-645.
- [4] Rizal, F., & Wijaya, A. (2021). Aplikasi Monitoring Jumlah Gizi Yang Dibutuhkan Ibu Hamil Untuk Mengurangi Resiko Gangguan Fisik dan Mental Janin. *Explore IT!: Jurnal Keilmuan dan Aplikasi Teknik Informatika*, 13(1), 1-5.
- [5] Saputra, P. Y., Subhi, D. H., & Winatama, F. Z. A. (2019). Implementasi Sentimen Analisis Komentar Channel Video Pelayanan Pemerintah Di Youtube Menggunakan Algoritma Naïve Bayes. *Jurnal Informatika Polinema*, 5(4), 209-213.
- [6] Kusuma, A., & Nugroho, A. (2021). Analisa Sentimen pada Twitter Terhadap Kenaikan Tarif Dasar Listrik Dengan Metode Naïve Bayes. *Jurnal Ilmiah Teknologi Informasi Asia*, 15(2), 137-146.
- [7] Simanjuntak, H. E., & Windarto, W. (2020). Analisa Data Mining Menggunakan Frequent Pattern Growth pada Data Transaksi Penjualan PT Mora Telematika Indonesia untuk Rekomendasi Strategi Pemasaran Produk Internet. *Jurnal Media Informatika Budidarma*, 4(4), 914-923. DOI: 10.30865/mib.v4i4.2300.
- [8] Astari, N. M. A. J., Divayana, D. G. H., & Indrawan, G. (2020). Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Dan Informatika (JSI)*, 15(1), 27-29. DOI: 10.30864/jsi.v15i1.332.
- [9] Rasenda, R., Lubis, H., & Ridwan, R. (2020). Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter. *Jurnal Media Informatika Budidarma*, 4(2), 369-376. DOI: 10.30865/mib.v4i2.2051.
- [10] Damuri, A., Riyanto, U., Rusdianto, H., & Aminudin, M. (2021). Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako. *JURIKOM (Jurnal Riset Komputer)*, 8(6), 219-225. DOI: 10.30865/jurikom.v8i6.3655.
- [11] Putro, H. F., Vlandari, R. T., & Saptomo, W. L. Y. (2020). Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan. *Jurnal Teknologi Informasi dan Komunikasi (TIKoSIN)*, 8(2).
- [12] Ginting, V. S., Kusriani, K., & Taufiq, E. (2020). Implementasi Algoritma C4. 5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python. *Inspiration: Jurnal Teknologi Informasi dan Komunikasi*, 10(1), 36-44.
- [13] Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, 5(2), 697-711.



- [14] Najjichah, H., Syukur, A., & Subagyo, H. (2019). Pengaruh text preprocessing dan kombinasinya pada peringkasan dokumen otomatis teks berbahasa Indonesia. *Jurnal Cyberku*, 15(1), 1-11.
- [15] Jonathan, M., & Nataliani, Y. (2022). Analisis Sentimen Penilaian Masyarakat Indonesia terhadap GeNose pada Komentar Youtube Menggunakan Metode Naïve Bayes. *d'CARTESLAN: Jurnal Matematika dan Aplikasi*, 11(1), 1-8.