

Analisis Sentimen Komentar Youtube Mengenai Keputusan Mahkamah Agung Terhadap batasan umur Pilkada Dengan Menggunakan Algoritma Naive Bayes dan SMOTE

Muhamad Rizvi Roshan^{1*}, Ramanda md², Adikara Alif Nurrahman³, Ayu Sri Oktaviani⁴, Ahmad Rizky⁵, Abdul Rahman⁶

¹²³⁴⁵Program Studi Informatika, Universitas Multi Data Palembang, Palembang, Indonesia
rizviroshan10@mhs.mdp.ac.id, ramandamd@mhs.mdp.ac.id, adikaraalf@mhs.mdp.ac.id,
ayuuoktaviani@iitii@mhs.mdp.ac.id, ahmadrizky@mhs.mdp.ac.id, arahman@mdp.ac.id

Abstrak

Penelitian ini menganalisis sentimen dalam komentar *YouTube* mengenai putusan Mahkamah Agung tentang pembatasan usia pemilihan menggunakan algoritma *Naive Bayes* dan pendekatan SMOTE (*Synthetic Minority Over-Sampling Technique*). Data dikumpulkan dari komentar *YouTube* dan kemudian diubah menjadi kumpulan data untuk diperiksa dan untuk *dataset* berjumlah 1596 data pada *dataset*. Menggunakan algoritma *Naive Bayes*, komentar dikategorikan menjadi sentimen positif, negatif, dan netral. Penerapan SMOTE digunakan untuk mengatasi ketidakseimbangan dalam representasi kelas dalam kumpulan data. Hasil penelitian menunjukkan bahwa model yang menjalani pelatihan menggunakan SMOTE mencapai tingkat akurasi 81%, menunjukkan presisi penting 0,85 dan *recall* 0,92 untuk sentimen negatif, bersama dengan skor F1 0,89. Sebaliknya, model yang tidak menggunakan SMOTE menunjukkan tingkat akurasi 80%, meskipun menampilkan kemampuan terbatas dalam mendeteksi sentimen positif. Penelitian ini menyoroti efektivitas penggunaan metodologi SMOTE dalam meningkatkan kinerja model *Naive Bayes* untuk analisis sentimen pada komentar *YouTube* terkait peraturan pemerintah. Temuan penelitian ini diharapkan dapat memberikan wawasan berharga bagi pembuat kebijakan dalam mengevaluasi dan memperbaiki regulasi yang ada.

Kata kunci: Analisis Sentimen, Naive Bayes, SMOTE, Youtube, Mahkamah Agung.

Abstract

This research analyzes sentiment in YouTube comments regarding the Supreme Court's ruling on voting age restrictions using the Naive Bayes algorithm and the SMOTE (Synthetic Minority Over-Sampling Technique) approach. Data was collected from YouTube comments and then converted into a dataset for examination and for the dataset amounted to 1596 data on the dataset. Using Naive Bayes algorithm, comments were categorized into positive, negative and neutral sentiments. The application of SMOTE was used to address the imbalance in class representation in the dataset. The results showed that the model that underwent training using SMOTE achieved an accuracy rate of 81%, exhibiting an important precision of 0.85 and recall of 0.92 for negative sentiments, along with an F1 score of 0.89. In contrast, the model that did not use SMOTE showed an accuracy rate of 80%, although it displayed limited ability in detecting positive sentiments. This research highlights the effectiveness of using the SMOTE methodology in improving the performance of Naive Bayes models for sentiment analysis on YouTube comments related to government regulations. The findings of this study are expected to provide valuable insights for policymakers in evaluating and improving existing regulations.

Keywords: Sentiment Analysis, Naive Bayes, SMOTE, Youtube, Supreme Court.

1. PENDAHULUAN

Indonesia merupakan negara yang menganut sistem politik demokrasi, yang ditandai dengan pelaksanaan pemilihan kepala daerah (pilkada) untuk memilih calon kepala daerah dan wakil kepala daerah. Sejak tahun 2015, pilkada dilaksanakan

secara serentak untuk mengganti kepala daerah dan wakil kepala daerah yang masa jabatannya selama 5 tahun telah habis, kebijakan pilkada serentak ini diresmikan pada tahun 2015, meskipun terdapat pro dan kontra terhadap pelaksanaan pilkada serentak pada tahun 2020, tidak semua daerah yang menyelenggarakan pilkada

serentak di tahun 2020 setuju dengan kebijakan pemerintah tersebut. Bahkan pemerintah sendiri ingin mengevaluasi kebijakan pilkada serentak, namun hal ini dibantah oleh Kementerian Dalam Negeri yang menyatakan bahwa evaluasi bukan berarti kembali ke pilkada melalui DPRD, Catatan menunjukkan bahwa pilkada langsung melalui DPRD memiliki dampak negatif bagi sistem sosial masyarakat Indonesia, termasuk tingginya potensi konflik di beberapa daerah seperti Papua dan Aceh [14].

Dalam sistem ketatanegaraan Indonesia, kekuasaan kehakiman merupakan kekuasaan yang independen, dijalankan oleh Mahkamah Agung dan badan peradilan di bawahnya serta Mahkamah Konstitusi, kekuasaan ini berfungsi untuk menyelenggarakan peradilan demi menegakkan hukum dan keadilan. Hal ini ditegaskan dalam Pasal 24 ayat (1), ayat (2), dan ayat (3) Undang-Undang Dasar Negara Republik Indonesia Tahun 1945 (UUD 1945) [12].

Menurut teori klasik dalam hukum yang dikenal sebagai teori *trias politica* yang dikemukakan oleh *Montesquieu*, kekuasaan negara dibagi menjadi tiga bidang: legislatif, eksekutif, dan yudikatif, kekuasaan yudikatif berfungsi untuk menjalankan peradilan atau kekuasaan kehakiman, kekuasaan kehakiman harus bersifat independen, mandiri, dan terpisah dari kekuasaan eksekutif dan legislatif, yang merupakan salah satu prinsip utama negara hukum. Independensi dan kemandirian kekuasaan yudikatif dimaksudkan agar lembaga ini dapat menjalankan fungsi peradilannya untuk menegakkan hukum dan keadilan tanpa pengaruh atau intervensi dari kekuasaan lainnya [3]. Mahkamah Agung (MA) dalam menyelesaikan kasus hukum yang dasar hukumnya tidak dijelaskan secara terperinci dalam perundang-undangan. demi menegakkan keadilan dan kebahagiaan manusia [6].

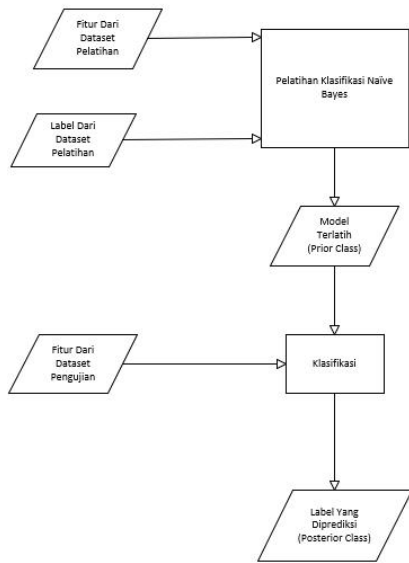
Pada penelitian jurnal yang terkait tentang analisis sentiment *youtube*. Pada penelitian [1] mengamati komentar *youtube*

tentang belajar *programming*. Penelitian ini dilakukan untuk menganalisis pandangan, penilaian, dan tanggapan pengguna terhadap konten pembelajaran pemrograman di *platform YouTube* dapat digunakan sebagai dasar untuk perbaikan atau pengembangan lebih lanjut dari konten edukasi tersebut. Dengan hasil akurasi dari *Naïve Bayes* menunjukkan bahwa prediksi untuk sentimen positif mencapai 100%, sedangkan prediksi untuk sentimen negatif adalah 66,39%.

Pada penelitian jurnal lainnya yang terkait tentang *naïve bayes* dan *SMOTE*. Pada penelitian [10] mengamati komentar pada *twitter* tentang *Complain* dan bukan *Complain*. Dari analisis jurnal ini, ditemukan bahwa 86,48% di antaranya adalah *Complain* dan 77,85% adalah bukan *Complain*. Dengan penerapan *SMOTE Upsampling*, jumlah sentimen komplain dan non-komplain yang benar menjadi seimbang, masing-masing berjumlah 581. Model ini mencapai tingkat akurasi rata-rata mikro sebesar 81.58%, presisi rata-rata mikro sebesar 86.48%, dan *recall* sebesar 74.87%, serta AUC sebesar 0.803. Penelitian ini diharapkan dapat membantu perusahaan untuk memahami umpan balik terhadap merek atau produk mereka.

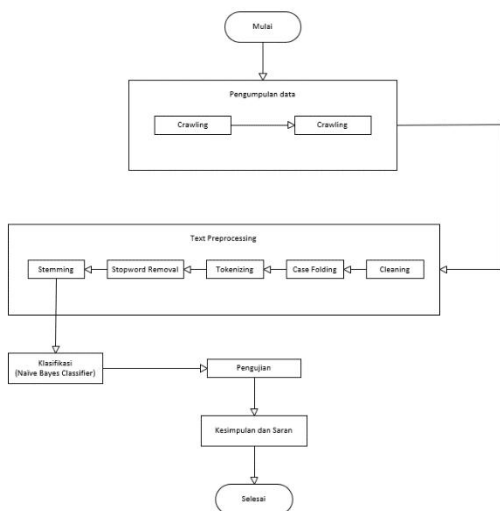
Penelitian ini membantu memahami persepsi publik dan sikap terhadap kebijakan pemerintah, terutama yang berkaitan dengan ambang usia untuk partisipasi pemilu. Wawasan tersebut berfungsi sebagai sumber daya yang signifikan bagi pengambil keputusan untuk menilai dan meningkatkan kebijakan ini.

2. METODE



Gambar 1. Diagram pelatihan dan identifikasi menggunakan klasifikasi *naive bayes*

Algoritma *Naive Bayes* adalah salah satu algoritma dalam teknik klasifikasi pada *data mining*. *Naive Bayes* menggunakan metode probabilitas dan statistik yang pertama kali diperkenalkan oleh ilmuwan Inggris, *Thomas Bayes*. Algoritma ini memprediksi kemungkinan kejadian di masa depan berdasarkan data dari masa lalu, sehingga dikenal sebagai *Teorema Bayes*[15].



Gambar 2. Tentang tahapan proses metodologi penelitian.

Menurut jurnal [9] *Naive Bayes* adalah metode yang tidak mengikuti aturan khusus dan menggunakan cabang matematika yang

dikenal sebagai teori probabilitas. Tujuannya adalah untuk mendapatkan kemungkinan tertinggi dengan menganalisis frekuensi atau jumlah kemunculan setiap klasifikasi dalam data pelatihan. *Naive Bayes* sering digunakan untuk mengklasifikasikan *dataset* berbasis teks. Keunikan dari *Naive Bayes* adalah asumsi sederhana (naif) dari probabilitas suatu kondisi atau peristiwa berdasarkan besarnya pengaruh [11].

Menurut *Charles Joergensen E. Munthe*, *Text Mining* adalah sebuah teknik yang dapat digunakan untuk klasifikasi. *Text mining* merupakan variasi dari *data mining* yang berusaha menemukan pola-pola yang menarik dalam kumpulan data teks yang besar. Selain itu, *text mining* juga dapat didefinisikan sebagai penerapan konsep-konsep dari teknik *data mining* untuk menemukan pola dalam teks, dengan tujuan mendapatkan informasi yang bermanfaat untuk tujuan tertentu [8].

Analisis sentimen itu proses otomatis untuk mengekstrak sikap, opini, dan emosi yang terkandung dalam data teks [7]. Analisis sentimen mengklasifikasikan perasaan berdasarkan polaritas teks dalam sebuah frasa, sehingga dapat diidentifikasi sebagai sentimen positif, negatif, atau netral [13].

Metode SMOTE (*Synthetic Minority Over-Sampling Technique*) adalah teknik yang digunakan untuk mengatasi masalah ketidakseimbangan kelas. Teknik ini mensintesis sampel baru dari kelas minoritas untuk menyeimbangkan *dataset* dengan membuat sampel baru dari kelas minoritas dengan membentuk cembung dari contoh-contoh yang berdekatan. Tingkat pengambilan sampel berlebih yang diperlukan kemudian dipilih secara acak. Teknik ini secara efektif menarik garis antara titik-titik minoritas dalam ruang fitur dan sampel di sepanjang garis ini. Contoh-contoh baru dari minoritas ini ditambahkan ke data pelatihan, dan pengklasifikasi dilatih dengan data tambahan. Algoritma SMOTE umumnya lebih akurat dibandingkan dengan pendekatan pengambilan sampel yang biasa [5].

Youtube sebagai salah satu situs media sosial yang menyediakan *video online* yang dapat diakses oleh siapa saja dan di mana saja selama ada koneksi internet [4].

3. HASIL DAN PEMBAHASAN

Untuk pengumpulan data sumbernya berasal dari komentar *youtube*,terus data tersebut diubah menjadi *dataset* menggunakan *netlytic* yang disimpan dengan file CSV dan untuk data pada *dataset* berjumlah 1596 data. Pada tabel 1 menunjukkan data dari komentar *youtube* yang telah diubah menjadi *dataset*.

Tabel 1. *Dataset*

author	description	pubdate
@hjwwetty1866	Hadeeeew skaesang emng udh mahir gtu di bidang peumerintahan...jd ketum jg ujug ujug pedah anak presiden....?? ?	2024-06-02 02:36:39
@wawanirawan6430	Negara kekuasaan hukum diutak atik utk kepentingan klrng	2024-06-02 02:33:33
@WayanApp	Ga usah rakut boss yg nentukan rakyat juga kenapa mesti baper	2024-06-02 02:20:02
@HendriDemokrasi	ERA SEKARANG AJA RAKYAT BANYAK YG JADI PENGECUT BISANYA NGOCEH DI	2024-06-02 02:19:48

	SOSMED BEDA ZAMAN ERA ORDE BARU	
@mudjionohadisap utro5665	Dah 17 tahun keatas aja biar banyak kesempatan generasi muda dan bukan generasi keluarga...	2024-06-02 02:17:51

Setelah *crawling data*, *dataset* akan melakukan proses yaitu *preprocessing Data*.tahap *preprocessing Data* meliputi pembersihan teks dengan mengubah huruf kecil menjadi konsistensi,normalisasi teks dengan mengubah kata-kata informal atau kata-kata singkat menggunakan Bahasa substitusi,dan penghapusan *stopwords* dengan menggunakan *sastrawi* untuk menghapuskan kata-kata umum yang tidak mempunyai makna penting.

Tabel 2. Hasil *Processing*

author	description	pubdate
@hjwwetty1866	hadeeeew skaesang emng udh mahir gtu di bidang peumerintahanjd ketum jg ujug ujug pedah anak presiden	2024-06-02 02:36:39
@wawanirawan6430	negara kekuasaan hukum diutak atik utk kepentingan klrng	2024-06-02 02:33:33
@WayanApp	ga usah rakut boss yg nentukan rakyat juga kenapa mesti baper	2024-06-02 02:20:02
@HendriDemokrasi	era sekarang aja rakyat banyak yg jadi pengecut bisanya ngoceh di sosmed beda zaman era orde baru	2024-06-02 02:19:48
@mudjionoha	dah 17 tahun keatas aja biar banyak	2024-06-02

akurasi yang sangat tinggi, sedangkan untuk kelas positif, itu menyusut menjadi hanya 0,10, menunjukkan ketidakmampuan yang signifikan. Selanjutnya, skor F1 untuk kelas negatif mencapai 0,89, sangat kontras dengan sedikit 0,16 yang dicapai untuk kelas positif, menggarisbawahi kemampuan model yang berkurang dalam membedakan sentimen positif tanpa adanya SMOTE. Rata-rata tertimbang yang dihitung untuk presisi, ingatan, dan skor F1 masing-masing ditentukan menjadi 0,75, 0,80, dan 0,74.

4. KESIMPULAN

Temuan penelitian ini menunjukkan bahwa pemanfaatan algoritma *Naive Bayes* di samping pendekatan SMOTE (*Synthetic Minority Over-Sampling Technique*) telah secara efektif meningkatkan kemanjuran analisis sentimen pada komentar *YouTube* yang berkaitan dengan putusan Mahkamah Agung tentang ambang usia untuk pemilihan. Hasil penilaian mengungkapkan bahwa penggabungan SMOTE menghasilkan akurasi model 81%, menampilkan presisi

yang sangat tinggi untuk sentimen negatif (0,85) dan *recall* (0,92), bersama dengan Skor F1 0,89. Namun demikian, peningkatan lebih lanjut diperlukan agar model dapat mengklasifikasikan sentimen positif secara akurat, seperti yang ditunjukkan oleh presisi 0,54, *recall* 0,36, dan Skor F1 0,43.

Sebaliknya, model yang tidak menggunakan SMOTE menunjukkan akurasi 80%, dengan presisi untuk sentimen negatif di 0,81 dan *recall* pada 0,98. Namun, model-model ini berjuang secara signifikan dalam mengidentifikasi sentimen positif, dengan tingkat *recall* hanya 0,10 dan Skor F1 0,16.

Intinya, penelitian menggaris bawahi efektivitas SMOTE dalam mengurangi ketidakseimbangan kelas dalam kumpulan data, akibatnya meningkatkan kemampuan model *Naive Bayes* untuk mengkategorikan sentimen negatif secara akurat. Studi ini menghasilkan wawasan berharga tentang opini publik mengenai kebijakan pemerintah dan menyoroti prospek yang menjanjikan terkait dengan memanfaatkan perpaduan algoritma *Naive Bayes* dan SMOTE untuk analisis sentimen berdasarkan data tekstual.

DAFTAR PUSTAKA

- [1] Adelia, S., Milanda, E., Santari, J., Kesuma, D. T., Silvia, E., & Kurniawan, F. (2023). Analisis Sentimen Belajar Programming Pada Media Sosial Youtube Menggunakan Algoritma Klasifikasi Naive Bayes. *Journal of Information Technology Ampera*, 4(3), 2774–2121. <https://journal-computing.org/index.php/journal-ita/index>
- [2] Agustina, D. A., Subanti, S., & Zukhronah, E. (2021). Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Marketplace di Indonesia Menggunakan Algoritma Support Vector Machine. *Indonesian Journal of Applied Statistics*, 3(2), 109. <https://doi.org/10.13057/ijas.v3i2.44337>
- [3] Dm, M. Y., Rani, S., Tabrani, S., Salwani, A., & Saragih, G. M. (2023). Politik Hukum Kewenangan Mahkamah Agung Dan Mahkamah Konstitusi di Indonesia. *Jurnal Pendidikan Dan Konseling*, 5, 698–705.
- [4] Firsttama, R. A., Arifiyanti, A. A., & Kartika, D. S. Y. (2024). Analisis Sentimen Komentar Youtube Konferensi Tingkat Tinggi G20 Menggunakan Metode Naive Bayes. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 6(2), 282–285. <https://doi.org/10.47233/jteksis.v6i2.1263>
- [5] Hermanto, Kuntoro, A. Y., Asra, T., Pratama, E. B., Effendi, L., & Ocanitra, R. (2020). Gojek and Grab User Sentiment Analysis on Google Play Using Naive Bayes Algorithm and Support Vector Machine Based Smote Technique. *Journal of Physics: Conference Series*, 1641(1).

- <https://doi.org/10.1088/1742-6596/1641/1/012102>
- [6] Islamiyati, I., Agus Priyono, E., Hendrawati, D., & Arief Budiman, A. (2020). Eksistensi Yurisprudensi Mahkamah Agung (MA) Dalam Penegakan Hukum Keluarga Islam Indonesia. *Law, Development and Justice Review*, 3(1), 1–14. <https://doi.org/10.14710/ldjr.v3i1.7349>
- [7] Isnain, A. R., Sulistiani, H., Hurohman, B. M., Nurkholis, A., & Styawati, S. (2022). Analisis Perbandingan Algoritma LSTM dan Naive Bayes untuk Analisis Sentimen. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 8(2), 299. <https://doi.org/10.26418/jp.v8i2.54704>
- [8] Karimah, A., Dwilestari, G., & Mulyawan, M. (2024). Analisis Sentimen Komentar Video Mobil Listrik Di Platform Youtube Dengan Metode Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 767–737. <https://doi.org/10.36040/jati.v8i1.8373>
- [9] Khoirul Insan, M. K., Hayati, U., & Nurdiawan, O. (2023). Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 478–483. <https://doi.org/10.36040/jati.v7i1.6373>
- [10] Kurniawan, B., Suwarisman, A., Afriyanti, I., Wahyudi, A., & Saputra, D. D. (2023). Analisis Sentimen Complain dan Bukan Complain pada Twitter Telkomsel dengan SMOTE dan Naive Bayes. *Jurnal JTIC (Jurnal Teknologi Informasi Dan Komunikasi)*, 7(1), 106–113. <https://doi.org/10.35870/jtik.v7i1.691>
- [11] Matondang, A. F., Dur, S., Cipta, H., Islam, U., & Sumatera Utara, N. (n.d.). *Proximal: Jurnal Penelitian Matematika dan Pendidikan Matematika ANALISIS SENTIMEN JASA EKSPEDISI PENGIRIMAN BARANG MENGGUNAKAN METODE NAIVE BAYES*. 7, 377–386. <https://doi.org/10.30605/proximal.v5i2.365>
- [12] Naibaho, R., & Hasibuan, I. J. M. (2021). Peranan Mahkamah Agung Dalam Penegakan Hukum Dan Keadilan Melalui Kekuasaan Kehakiman. *Nommensen Journal of Legal Opinion*, 2(02), 203–214. <https://doi.org/10.51622/njlo.v2i02.388>
- [13] Ningsih, W., Alfianda, B., Rahmaddeni, R., & Wulandari, D. (2024). Perbandingan Algoritma SVM dan Naive Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 556–562. <https://doi.org/10.57152/malcom.v4i2.125>
- [14] Novrisal, A., Marthasari, G. I., & Aditya, C. (2021). Sentimen Analisis Tweet Berbahasa Indonesia Pada Pilkada Serentak 2020 Menggunakan Metode Naive Bayes Berbasis Particle Swarm Optimization. *Jurnal Repositor*, 3(2), 191–198. <https://doi.org/10.22219/repositor.v3i2.1169>
- [15] Nurmayanti, W. P. (2021). Penerapan Naive Bayes dalam Mengklasifikasikan Masyarakat Miskin di Desa Lepak. *Geodika: Jurnal Kajian Ilmu Dan Pendidikan Geografi*, 5(1), 123–132. <https://doi.org/10.29408/geodika.v5i1.3430>
- [16] Chely Aulia Misrun, Haerani, E., Fikry, M., & Budianita, E. (2023). Analisis sentimen komentar youtube terhadap Anies Baswedan sebagai bakal calon presiden 2024 menggunakan metode naive bayes classifier. *Jurnal CoSciTech (Computer Science and Information Technology)*, 4(1), 207–215.

<https://doi.org/10.37859/coscitech.v4i1.4790>

- [17] Verma, D., Agarwal, H., & Aggarwal, A. K. (2019). *Recognition of*

Degraded and Non Degraded Roman Characters Using Different Classifiers. 6(5), 402–419.
<https://doi.org/10.13189/ujeee.2019.060510>