Journal of Information Technology Ampera
Vol. 4, No. 3, 2023 e-ISSN: 2774-2121
https://journal-computing.org/index.php/journal-ita/index

Analisis Sentimen Persepsi Publik Terhadap Kasus *Bullying* Siswa Cilacap Menggunakan Pendekatan *Machine Learning*

Muhammad Alfarizi[1], Muhammad Rizqy[2], Rifqi Imam Ghufroni[3], Dzaki Fathurahman[4], Rahmad Dani Saputra[5], Fandi Kurniawan[6]

[1,4,5,6]Sistem Informasi, Universitas Muhammadiyah Kotabumi, Kotabumi, Indonesia
[2,3] Sistem Teknologi Informasi, Universitas Muhammadiyah, Kotabumi, Indonesia
Email: [1]Malfa.2059201079@umko.ac.id, [2]muham.2059201009@umko.ac.id,
[3] rifqi.2059201016@umko.ac.id, [4]dzaki.2059201023@umko.ac.id,
[5]rahmatdani46970@gmail.com, [6]fandi.kurniawan@umko.ac.id

Abstrak

Analisis sentimen jadi pendekatan yang berarti dalam memahami anggapan dan opini masyarakat terhadap suatu peristiwa, sangat utama kala mengaitkan aksi kekerasan semacam bullying. Kasus bullying siswa SMP di Cilacap, Jawa Tengah, pada bulan September 2023 jadi sorotan publik dan menarik atensi melalui media sosial, sangat utama di platform YouTube. Studi ini dicoba dengan tujuan utama mempraktikkan algoritma klasifikasi Naive Bayes dalam melakukan analisis sentimen terhadap komentar di media sosial terpaut kasus bullying. Tata metode studi mengaitkan crawling data, pra-pemrosesan mengenakan RapidMiner, konsumsi TF- IDF buat representasi dokumen, dan implementasi Naive Bayes Classifier. Hasil analisis sentimen menunjukkan tingkatan akurasi sebesar 98. 19%, dengan kemampuan yang besar dalam mengklasifikasikan komentar sebagai positif, negatif, maupun netral. Meski demikian, evaluasi mengenakan confusion matrix berkata sebagian kasus negatif yang teridentifikasi tidak benar. Kesimpulan dari studi ini dapat memberikan pengetahuan berharga tentang tata cara masyarakat merespons dan berpendapat terhadap aksi bullying di zona sekolah.

Kata Kunci: Analisis Sentimen, Bullying Siswa SMP, Media Sosial, YouTube, Algoritma Naive Bayes.

I. PENDAHULUAN

Di era globalisasi dan kemajuan teknologi yang pesat, masalah kekerasan, terutama perundungan terhadap siswa, telah menjadi sorotan utama di masyarakat. Perundungan bukan lagi sekadar fenomena lokal; ia telah menyeberang batas geografis, memasuki dunia media sosial, dan menimbulkan tantangan baru dalam penanganan dan pencegahannya [1]. Contoh nyata dari masalah ini terjadi pada September 2023, di mana insiden kekerasan di sebuah SMA di Cilacap, Jawa Tengah, mendapat perhatian luas setelah video kejadian tersebut tersebar di media sosial. Video tersebut menampilkan siswa yang

terlibat masih mengenakan seragam sekolah, menimbulkan kekhawatiran dan diskusi publik mengenai fenomena perundungan di kalangan pelajar [2].

Kasus ini tidak hanya menarik perhatian terhadap masalah perundungan di sekolah, tetapi juga memicu debat mengenai respon publik terhadap kebijakan pemerintah, khususnya yang berkaitan dengan isu perpindahan ibu kota negara. Di sinilah YouTube, sebagai salah satu platform media sosial utama, berperan penting dalam menggambarkan pandangan masyarakat. Teknik data mining menjadi alat krusial dalam proses ini, memungkinkan pengklasifikasian, prediksi, dan pengelompokan data untuk memahami opini publik secara lebih mendalam [3].

Tujuan utama dari penelitian ini adalah untuk mengidentifikasi dan menganalisis emosi warga dalam respon mereka terhadap isu perundungan, dengan fokus khusus pada interaksi di media sosial. Analisis sentimen akan digunakan untuk membedakan tanggapan masyarakat yang negatif, positif, dan netral, memberikan gambaran yang lebih jelas mengenai pandangan dan sikap mereka terhadap isu ini. Pemahaman yang lebih detil terhadap respon ini diharapkan bisa memberikan kontribusi yang signifikan terhadap pengembangan strategi pencegahan perundungan yang lebih efektif dan menyeluruh [4].

Dalam rangka mencapai tujuan ini, penelitian akan mengadopsi metode analisis benih untuk mengungkap pola sikap dan pendapat masyarakat. Pendekatan ini akan memperluas pemahaman tentang sikap masyarakat terhadap perundungan, dengan mengumpulkan dan menganalisis data dari berbagai sumber, termasuk video dari YouTube yang telah menarik perhatian publik. Melalui pendekatan ini, penelitian berupaya untuk memberikan pandangan yang lebih komprehensif tentang situasi perundungan di kalangan siswa SMA Cilacap.

Penelitian ini akan menggunakan algoritma klasifikasi Naive Bayesian dalam analisis sentimen untuk mencapai keakuratan yang lebih tinggi dalam mengklasifikasikan respon warga. Dengan demikian, hasil penelitian ini diharapkan dapat memberikan wawasan yang berharga dalam menanggapi dan menangani masalah perundungan di kalangan pelajar, serta menginformasikan strategi pencegahan yang lebih efektif dan berkelanjutan untuk masa depan.

2. METODE

2.1 Algoritma *Naive Bayes*

Algoritma Naive Bayes Classifier ialah algoritma yang digunakan untuk mencari nilai probabilitas paling tinggi dan mengklasifikasi informasi uji pada jenis yang

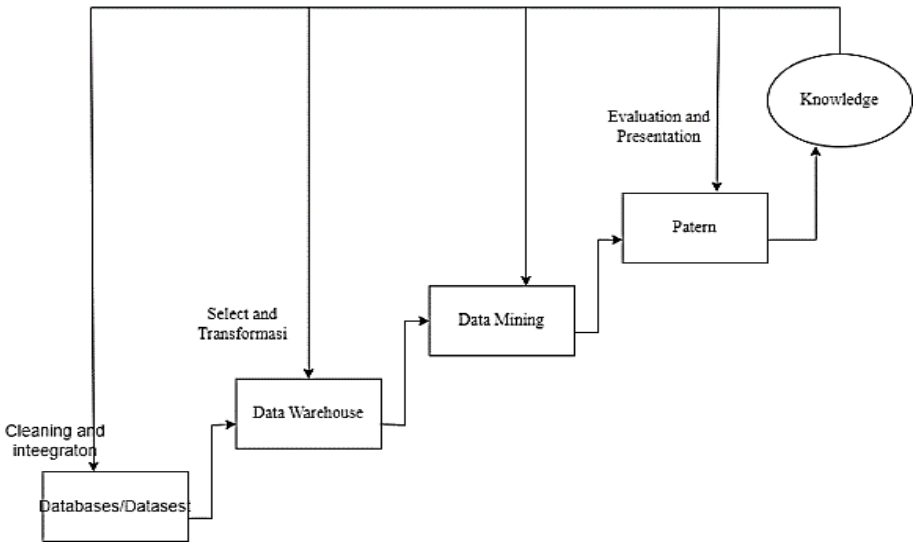
sangat pas di masa depan. Pada metode ini dianggap mudah dan efektif di gunakan pada analisis perusahaan. Naïve bayes menggunaka tenik supervised klasifikasi objek untuk masa depan dengan menerapkan label kelas atau catatan menggunakn probabilitas bersyarat[3]. Berikut beberapa komponen utama Naive Bayes.

1) Teorema Bayes: Naive Bayes didasarkan pada teorema Bayes, yang menyatakan hubungan antara distribusi probabilitas bersyarat. Untuk kejadian *A* dan *B*, teorema Bayes dapat dituliskan sebagai: $P(y|X) = P(X)\,P(X|y) \times P(y)$
2) Kemerdekaan: Secara khusus, Naive Bayes berasumsi bahwa semua fitur yang digunakan untuk mendeskripsikan suatu kasus atau objek tidak bergantung satu sama lain karena kelasnya. Meskipun asumsi-asumsi ini seringkali tidak sesuai dengan situasi nyata, keunggulan komputasi yang diberikannya menjadikannya metode yang cepat dan efisien.
3) Evaluasi: Naive Bayes digunakan untuk masalah klasifikasi dimana kita mencoba memprediksi kelas suatu kasus berdasarkan karakteristiknya. Ini dapat digunakan, misalnya, dalam klasifikasi spam email, diagnosis medis, atau klasifikasi dokumen.
4) Belajar dari data: Model Bayesian yang naif belajar dari data pelatihan yang diberikan. Selama klasifikasi, model ini mempelajari distribusi probabilitas fitur di setiap kelas. Dengan kata lain, model ini mempelajari seberapa sering suatu fitur muncul di setiap kelas.
5) Rumus dasar: Dalam klasifikasi biner, Naive Bayes menggunakan rumus dasar: $P(y|X) = P(X)P(X|y) \cdot P(y)$ Di mana: $P(y|X)$adalah probabilitas kelas *y* jika diberikan properti *X*. $P(X|y)$adalah probabilitas properti *X* dari kelas tertentu *y*. $P(y)$adalah probabilitas prior dari kelas *y*. $P(X)$adalah probabilitas prior dari fitur *X*.
6) Aplikasi: Bergantung pada distribusi data dan asumsi yang relevan, Naive Bayes dapat digunakan dalam beberapa variasi, termasuk Naive Bayes Gaussian, Naive Bayes Multinomial, dan Naive Bayes Bernoulli. Meskipun asumsi independensi yang kuat menjadikannya "naif", namun Naive Bayes seringkali memberikan hasil yang baik dalam praktiknya dan sering digunakan dalam berbagai aplikasi klasifikasi.

2.2 *Data Mining*

Informasi mining merupakan serangkaian proses buat mengekplorasi informasi memakai pc dengan proses ekstraksi serta bisa menggali pola berarti informasi. Informasi mining dipecah jadi sebagian kelompok ialah: Prediksi, Ekstimasi, Klasifikasi, Pengklusteran serta Asosia. Informasi mining selaku Metode pendidikan pc mempunyai sebagian tahapan supaya menciptakan pengetahuan.

Berikut ini ialah proses pengolahan informasi mining [6]. Data mining dapat didefinisikan sebagai suatu proses untuk dapat mencari pola dari beberapa kumpulan data yang terdapat didalam database yang kemudian dianalisis sehingga dapat menghasilkan suatu informasi untuk dimanfaatkan pada proses berikutnya[10]. Gambar 1 merupakan tahapan data *mining*.



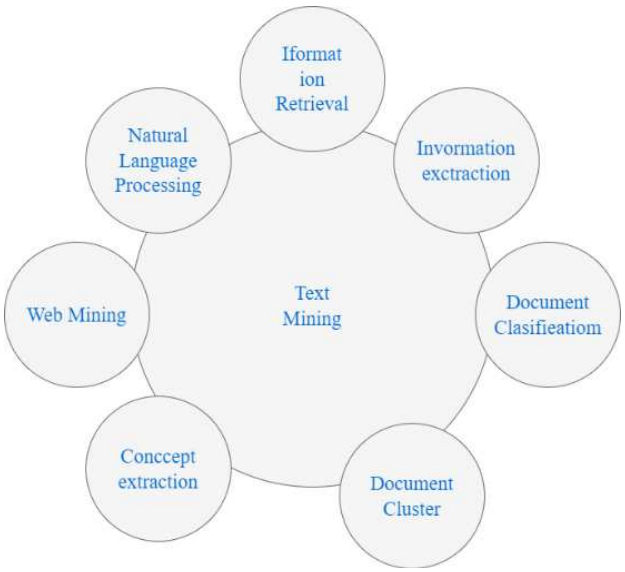
Gambar 1. Tahapan *Data Mining*

Berikut ini tahapan-tahapan Knowledge Mining atau Penemuan Pengetahuan dalam Basis Data (KDD).

1) Integrasi Data: Ini adalah tahap awal di mana sampel data dari berbagai sumber dikumpulkan dan digabungkan.
2) Pemilihan data: Langkah ini terdiri dari pembuatan kumpulan data target atau pemfokusan pada subkumpulan variabel dan pola informasi yang hasilnya akan diuji.
3) Pembersihan data: Proses pembersihan terlebih dahulu data yang tersimpan untuk mendapatkan data yang tidak berubah.
4) Transformasi Data: Sesi untuk mengubah data menjadi informasi yang valid.
5) Penambangan data: Langkah ini terdiri dari pencarian pola-pola menarik dalam representasi yang diberikan tergantung pada tujuan penambangan data.
6) Evaluasi: Langkah ini terdiri dari interpretasi dan evaluasi pola yang ditambang.
7) Informasi: Langkah terakhir dalam membantu pengguna mengambil keputusan berdasarkan informasi yang mereka terima.

2.3 *Text Mining*

Text mining merupakan proses mengekstrak data dari kumpulan dokumen dengan memakai tools analisis yang ialah komponen dari informasi mining. Tools analisis tersebut bisa digunakan buat kategorisasi data ataupun bacaan, dan pengelompokan bacaan. Sumber informasi yang digunakan dalam text mining merupakan kumpulan pola bahasa natural yang sanggup menganalisis informasi bacaan semi- terstruktur serta tidak terstruktur. Analisis sentimen merupakan salah satu cabang ilmu text mining yang menekuni metode buat mencerna serta menganalisis emosi yang tercantum dalam sesuatu bacaan. Tujuan dari analisis sentimen merupakan buat mengklasifikasikan sentimen pada sesuatu bacaan, baik itu sentimen positif, negatif, ataupun netral.[7]. *Text mining* bisa dipecah jadi 7 bidang bersumber pada karakteristiknya. Walaupun mempunyai perbandingan bidang, ketujuh bidang tersebut mempunyai keterkaitan satu sama lain. semacam yang tertera pada Gambar 2.



Gambar 2. 7 Bidang *Text Mining*

2.4 *Text Preprocessing*

Preprocessing merupakan proses buat mensterilkan serta mempersiapkan informasi mentah buat dianalisis. Proses ini umumnya dicoba dengan metode menghapus informasi yang tidak relevan ataupun mengganti informasi supaya lebih gampang dianalisis. Preprocessing sangat berarti dalam analisis sentimen, paling utama di media sosial, sebab bacaan di media sosial kerap kali informal,

tidak terstruktur, serta berisi *noise* [2]. *Pre-Processing Text* memiliki tahapan berikut.

1) *Cleaning* ataupun Pembersihan bacaan merupakan proses buat melenyapkan ciri baca serta kepribadian yang tidak relevan dari bacaan. Proses ini dicoba buat membuat bacaan lebih gampang dianalisis serta tingkatkan akurasi hasil analisis.
2) *Transform case*, Proses dimna mengganti seluruh huruf pada bacaan jadi huruf kecil. Transformasi huruf dicoba buat mempermudah analisis bacaan.
3) *Tokenization,* merupakan proses memecah bacaan jadi unit- unit kecil yang lebih gampang dianalisis.
4) *Stopwords* merupakan perkata universal yang tidak membagikan data berarti dalam analisis bacaan. Stopword umumnya digunakan dalam mesin pencari buat melenyapkan perkata yang tidak relevan dengan pencarian.

2.5 Analisis Sentimen

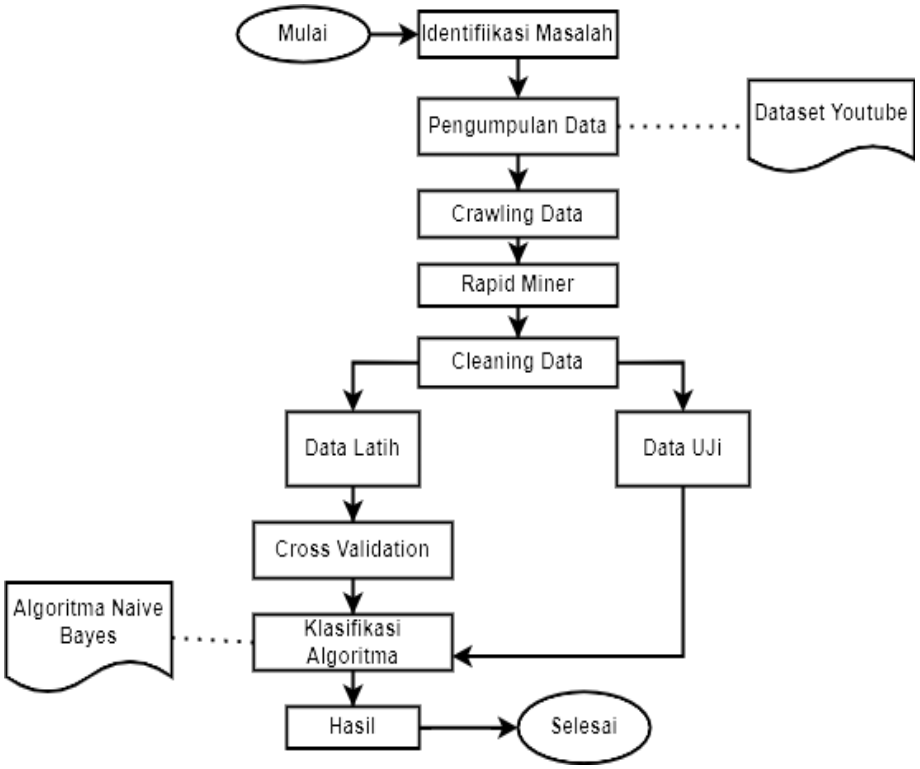
Analisis sentimen merupakan proses buat menguasai emosi yang tercantum dalam bacaan. Analisis sentimen bisa digunakan buat memastikan apakah bacaan tersebut mempunyai emosi netral, negatif ataupun positif.[5]. YouTube merupakan platform media sosial yang membolehkan pengguna buat menyaksikan ataupun mengunggah dan memberikan video secara free. YouTube bisa diakses oleh siapa saja di dunia, serta pada Oktober 2019, YouTube merupakan platform media sosial kedua yang sangat terkenal di dunia, dengan jumlah pengguna aktif menggapai 2 miliyar orang.

Naive Bayes Naive Bayes merupakan tata cara klasifikasi yang simpel serta kilat buat memprediksi kelas sesuatu ilustrasi. Tata cara ini memakai teorema Bayes buat menghitung probabilitas kemunculan fitur- fitur tertentu dalam sesuatu kelas. Misalnya, buat memprediksi apakah seorang mempunyai pekerjaan bergaji besar, kita bisa memakai tata cara Naive Bayes dengan fitur- fitur semacam umur, tipe kelamin, serta pemasukan. Bila kita mempunyai informasi tentang ilustrasi orang- orang yang mempunyai pekerjaan bergaji besar, kita bisa menghitung probabilitas timbulnya tiap fitur dalam kelas tersebut. Setelah itu, kita bisa memakai probabilitas ini buat memprediksi apakah seorang dengan fitur- fitur tertentu cenderung mempunyai pekerjaan bergaji besar.

3. HASIL DAN PEMBAHASAN

3.1 Tahapan Pelaksanaan

Alur dalam tahapan pelaksanaan analisis dilakukan dalam beberapa tahapan seperti Gambar 3.



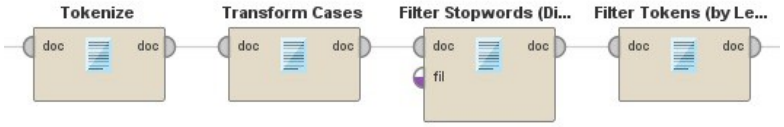
Gambar 3. Tahapan Pelaksanaan

3.2 *Crawling* Data

Dalam proses pengindeksan, diperoleh total 1175 komentar dengan menggunakan Google Colab Python. Peneliti membersihkan data dari nilai yang hilang, redundansi, dan duplikat. Metode crawling digunakan ketika pengumpulan informasi tersebut crawling merupakan cara untuk mendapatkan sebuah konten informasi yang ada pada web [8]. Awal mula digunakan nya Metode pengindeksan yang digunakan oleh browser adalah dengan mengisi indeksnya. Pengindeksan struktur data yang besar dikaitkan dengan kemacetan dan kurangnya manajemen konten [9].

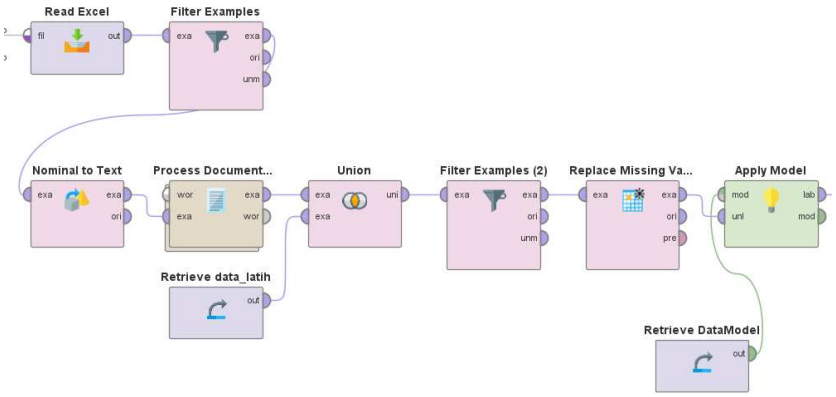
3.3 *Pre-Processing*

Pre-processing Data Pada proses *pre-processing*, data mentah disiapkan menjadi data set melalui proses pengolahan data untuk digunakan sebagai model dalam pembelajaran mesin. Adapun tahapan dari *pre-processing* data yaitu:



Gambar 4. Tahapan *Pre Processing* Data

Tahapan pre- processing informasi ialah diawali dari proses membaca file excel yang lebih dahulu telah ditaruh kala di crawling, setelah itu jalani pembersihan informasi(cleansing) yang ditaruh di dalam satu posisi spesial dinamakan subprocess, setelah itu dilanjutkan dengan memakai operator filter examples buat melenyapkan ataupun menghapus kolom informasi youtube yang kosong, tidak hanya itu pakai pula operator remove duplicates buat menghapus informasi ganda ataupun informasi youtube yang sama ataupun dicetak kesekian. Sehabis seluruh proses berakhir, informasi ditaruh ke dalam file excel yang baru selaku dataset buat proses analisis sentiment [1]. Proses berikutnya dalam *pre-processing* merupakan melaksanakan stemming ataupun dengan kata lain melenyapkan imbuhan di tiap perkata yang sudah diproses lebih dahulu sehingga keluaran dari proses pre- processing ini telah bisa digunakan buat dicoba perhitungan dengan memakai algoritma naive bayes classfier ataupun support vector machine. Buat proses stemming pada riset ini memakai website aplikasi secara online di web gataframework. com setelah itu ditambahkan dengan stemming memakai file *regexp.*



Gambar 5. Rangkaian Proses Pada *Rapidminer*

3.4 TF-IDF

Dataset yang sudah melewati proses pelabelan wajib berupa angka. Dataset tersebut dirubah jadi angka memakai tata cara TF- IDF. Tata cara ini bekerja buat memastikan keterikatan kata(term) terhadap dokumen dengan membagikan nilai pada tiap kata [4].

Row No.	aaammiinn	aamiin	aamiinbravo	aamiinn	aammiin
1	0	0.181	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
6	0	0	0	0	0
7	0	0	0	0	0
8	0	0	0	0	0
9	0	0	0	0	0
10	0	0	0	0	0
11	0	0	0	0	0
12	0	0	0	0	0
13	0	0	0	0	0
14	0	0	0	0	0
15	0	0	0	0	0

Gambar 6. Hasil TF=IDF

Row No.	word	in documents	total
1	negatif	271	271
2	anak	184	245
3	positif	223	223
4	pelaku	90	104
5	hukum	92	97
6	orang	83	96
7	biar	68	79
8	sekolah	54	75
9	polisi	61	73
10	korban	43	53
11	umur	43	52
12	penjara	35	38
13	kalo	29	34
14	guru	25	33

Gambar 7. Hasil Total TF-IDF

Pada Gambar 7 terdapat 14 *Row no*, pada *word* negatif, *in documents* 271 dan total 271. *Word* anak, *in document* 184 dan total 245. *Word* positif, *in documents* 223 dan total 223. *Word* pelaku, *in documents* 90 dan total 104. *Word* hukum, *in documents* 92 dan total 97. *Word* orang, *in documents* 83 dan total 96. *Word* biar, *in documents* 68 dan total 79. *Word* sekolah, *in documents* 54 dan total 75. *Word* polisi, *in documents* 61 dan total 73, *Word* korban, *in documents* 43 dan total 53. *Word* umur, *in documents* 43 dan total 52. *Word* penjara, *in documents* 35 dan total 38. W*ord* kalo, *in documents* 29 dan total 34. *Word* guru, *in documents* 25 dan total 33.

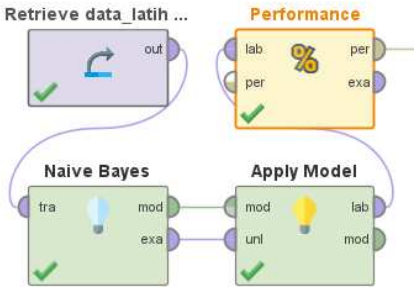
3.5 *WordCloud*

Word cloud merupakan visualisasi informasi bacaan yang menunjukkan perkata dari kumpulan informasi bacaan di sesuatu model yang telah dibentuk. Visualisasi tersebut menarangkan kalau terus menjadi besar dimensi sesuatu kata hingga perkata tersebut yang frekuensi kemunculannya lebih banyak ataupun kerap timbul dari sesuatu informasi bacaan [3].

Gambar 8. *Word Cloud* Hasil Analisa

3.6 *Naïve Bayes Classifier*

Dalam penelitian ini menggunakna salah satu algoritma Naïve Bayes. Ini telah biasa memandang tingkatan akurasi algoritma buat melaksanakan proses analisis sentiment. Dengan dorongan aplikasi rapidminer digunakan operator Naïve Bayes Classifier buat tersebut percobaan semacam foto dibawah ini[5].



Gambar 9. *Naïve Bayes Classifier Operator*

Pada algoritma *Naïve Bayes Classifier* dataset yang digunakan berjumlah 1175 total keseluruhan yang sudah dibersihkan atau difilter pada proses sebelumnya, selanjutnya data dibagi menjadi 74% data latih dan 26% data uji dengan pembagian data latih yang digunakan 875 yang sudah diberi sentiment dan data uji 298 yang tidak diberi sentiment. Hasil nilai akurasinya berada pada angka 93,71% seperti Gambar 10.

accuracy: 98.19%

	true negatif	true positif	class precision
pred. negatif	269	9	96.76%
pred. positif	0	220	100.00%
class recall	100.00%	96.07%	

Gambar 10. Akurasi *Naïve Bayes Classifier*

4. KESIMPULAN

Analisis dari hasil penelitian ini mengungkapkan bahwa sistem atau model yang digunakan memiliki presisi yang sangat tinggi, yaitu 100%, dalam mengklasifikasikan respon positif. Ini menunjukkan bahwa semua prediksi positif yang dilakukan oleh sistem adalah akurat. Namun, tingkat recall untuk kelas positif adalah 92,64%, yang menandakan bahwa ada sejumlah kasus positif yang belum teridentifikasi sepenuhnya oleh sistem. Sementara itu, untuk kelas negatif, presisi yang dicapai adalah 69,95%, yang mengindikasikan bahwa terdapat kemungkinan kecil beberapa prediksi negatif mungkin tidak akurat. Namun, tingkat recall untuk kelas negatif mencapai 100%, artinya sistem berhasil mengidentifikasi semua kasus negatif dengan benar. Kesimpulan ini memberikan wawasan penting mengenai efektivitas dan keterbatasan sistem dalam mengklasifikasikan berbagai jenis respon.

REFERENSI

[1] A. Tan, "Text Mining: The state of the art and the challenges," Bulletin of Roszdravnadzor, vol. 4, pp. 9–15, 2017.

[2] Y. Saraswati, T. Suprihatiningsih, and S. Pranowo, "Pengaruh Pendidikan Kesehatan Tentang Bullying Dengan Metode Ceramah Menggunakan Leaflet Dan Lcd Terhadap Sikap Bullying Pelajar Smpn 4 Cilacap," Pros. Semin. Nas. dan Disem. Penelit. Kesehat., vol. 1, no. 1, pp. 125–128, 2018.

[3] S. Sahar, "Analisis Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Clasiffier Pada Dataset Penyakit Jantung," Indones. J. Data Sci., vol. 1, no. 3, pp. 79–86, 2020, doi: 10.33096/ijodas.v1i3.20.

[4] A. Deolika, K. Kusrini, and E. T. Luthfi, "Analisis Pembobotan Kata Pada Klasifikasi Text Mining," J. Teknol. Inf., vol. 3, no. 2, p. 179, 2019, doi: 10.36294/jurti.v3i2.1077.
[5] M. Christianto, J. Andjarwirawan, and A. Tjondrowiguno, "Aplikasi analisa sentimen pada komentar berbahasa Indonesia dalam objek video di website YouTube menggunakan metode Naïve Bayes classifier," J. Infra, vol. 8.1, pp. 255–259, 2020.
[6] J. Budiarto, "Identifikasi Kebutuhan Masyarakat Nusa Tenggara Barat pada Pandemi Covid-19 di Media Sosial dengan Metode Crawling," vol. 2, no. 4, pp. 244–250, 2021.
[7] A. Putri dan A. Muzakir, "Analisis sentimen cyberbullying kpop di media sosial twitter menggunakan metode naive bayes," 2022.
[8] Calyptra, "Studi deskriptif perilaku bullying pada remaja," 2014.
[9] D. Krstinić, M. Braović, L. Šerić, dan D. Božić-Štulić, "Multi-label classifier performance evaluation with confusion matrix," 2020.
[10] G. L. Webb, E. Keogh, dan R. Miikkulainen, "Naïve Bayes," 2010.
[11] L. Ardiani, H. Sujaini, dan T. Tursina, "Implementasi sentiment analysis tanggapan masyarakat terhadap pembangunan di Kota Pontianak," 2020.
[12] Y. Liu, X. Huang, A. An, dan X. Yu, "SSAP: Storylines and sentiment aware pre-trained model for story ending generation," 2007.
[13] S. Raghavan dan H. Garcia-Molina, "Crawling the Hidden Web," 2000.
[14] M. Najork dan A. Heydon, "High-performance website crawling," pp. 25–45, Springer US, 2002.
[15] B. Mirkin, "Informasi analysis, mathematical statistics, machine learning, informasi mining: Similarities and differences," in 2011 International Conference on Advanced Computer Science and Information Systems, pp. 8, December 2011.