

**Algoritma Machine Learning Naïve Bayes pada Analisis Sentimen
Kesepakatan Polri dan GNPf-MUI pada Aksi Bela Islam III '212'**

**Angga Dri Hananto¹, Alivia Measy Erfiana², Bella Lexiani
Permata Putri³, Pelita Dwi Putri⁴, Fandi Kurniawan⁵**

Universitas Muhammadiyah Kotabumi¹²³⁴⁵

E-mail: angga.2059201090@umko.ac.id¹,
alivi.2059201106@umko.ac.id², bella.2059201082@umko.ac.id³,
pelit.2059201054@umko.ac.id⁴, fandi.kurniawan@umko.ac.id⁵

Abstrak

Indonesia adalah negara dengan masyarakat yang beragam, membentang dari Sabang hingga Merauke, di mana terdapat berbagai suku, agama, budaya, dan adat istiadat. Dengan keberagaman ini, masyarakat Indonesia diharapkan memiliki sikap toleransi yang tinggi. Data mining adalah proses mencari pola atau informasi dari data yang dipilih menggunakan teknik atau metode tertentu. Metode dalam data mining bervariasi, pemilihan algoritma yang tepat sangat bergantung pada hasil dan tujuan Knowledge Discovery in Databases (KDD) secara keseluruhan. Tahapan dalam proyek ini meliputi: pengambilan dataset, preprocessing dataset, TF-IDF, algoritma, kinerja. Algoritma yang digunakan dalam analisis ini adalah: Naïve Bayes. Algoritma Naïve Bayes merupakan metode klasifikasi yang berakar pada teorema Bayes. Ia memiliki karakteristik utama dengan asumsi yang sangat kuat tentang kondisi yang terjadi.

Kata Kunci — Indonesia, Data Mining, Naïve Bayes.

PENDAHULUAN

Indonesia merupakan Negara dengan keanekaragaman masyarakatnya yang membentang dari Sabang hingga Merauke, dimana banyaknya suku, agama, budaya, kebiasaan yang beraneka ragam. Dengan adanya hal tersebut membuat masyarakat Indonesia harus memiliki sikap toleransi yang tinggi. Toleransi ialah sikap saling menghargai dan menghormati satu dengan yang lain. Indonesia merupakan negara yang menghargai berbagai macam perbedaan, yang tertuang dalam Bhineka Tunggal Ika. Selain itu ini tercermin dalam dasar negara yaitu Pancasila, yang mengakui dan mendorong toleransi salah satunya antar agama yang ada di Indonesia.

Pancasila sebagai ideologi dasar negara Indonesia menekankan pada sila kelima yang berbunyi “Keadilan Sosial Bagi Seluruh Rakyat Indonesia”, yang mencakup prinsip-prinsip kesetaraan dan perlindungan hak agama bagi semua warga negaranya. Dalam praktiknya, Indonesia telah menjadi rumah bagi berbagai komunitas agama, termasuk Islam, Kristen, Hindu, Budha, dan kepercayaan tradisional. Pancasila dan semangat toleransi telah memainkan

peran penting dalam membentuk kerukunan antar agama di Indonesia. Indonesia terus berusaha untuk mempromosikan toleransi, keragaman, dan kerukunan antar agama sebagai landasan penting bagi persatuan dan kemajuan bangsa.

Dalam Upaya mempertahankan tradisi toleransi agama, meskipun tantangan-tantangan tetap muncul dalam beberapa kasus, salah satu kasus yang menjadi tantangan bagi Indonesia dalam mempertahankan toleransi pada keanekaragaman salah satunya agama adalah “Aksi bela Islam III” atau sering disebut juga sebagai aksi 212, adalah sebuah gerakan massa yang signifikan terjadi di Indonesia. Gerakan ini mencapai puncaknya pada tanggal 2 Desember 2016, Ketika jutaan demonstran yang sebagian besar berasal dari kelompok islam konservatif berkumpul di Jakarta untuk menyuarakan ketidakpuasan mereka terhadap pemerintah.

Aksi ini diawali sebagai bentuk protes masyarakat terhadap dugaan penistaan agama yang dilakukan oleh Gubernur Jakarta saat itu, yaitu Basuki Tjahaja Purnama atau yang kerap dikenal sebagai Ahok. Demonstrasi ini menjadi sorotan nasional maupun internasional, menggambarkan ketegangan antara kelompok islam konservatif dan otoritas pemerintah hingga dilakukannya kesepakatan antara Polri dan GNPF-MUI untuk Aksi tersebut yang berlangsung di kantor MUI (Majelis Ulama Indonesia) sebagai bentuk rencana aksi damai. Kesepakatan tersebut dihadiri oleh jajaran polri salah satunya diwakili oleh Jenderal Polisis Tito Karnavian sebagai aparat keamanan dan ketua MUI serta ulama-ulama setempat.

Pertemuan ini membahas mengenai aksi Desember 2016, pada aksi ini demonstran akan melakukan kegiatan dzikir, tauziah, dan sholat jumat yang berlokasi di bundaran HI, tamrin, dan sudirman. Pihak kepolisian menyampaikan bahwa sesuai dengan undang-undang 9 tahun 1998 dalam pasal 6 yang mengatur bahwa “pelaksanaan unjuk rasa tidak diperkenankan mengganggu ketertiban umum dan mengganggu hak asasi masyarakat lainnya”, kemudian dalam pasal 15 diatur “apabila pasal 6 dilanggar oleh sekelompok masyarakat, maka aparat keamanan setempat dapat melakukan pembubaran sebagai bentuk penertiban masyarakat. Penertiban masyarakat ini dilakukan untuk mencegah unjuk rasa serupa yang dilakukan kelompok atau elemen-elemen tertentu masyarakat dikemudian hari.

Dalam pertemuan tersebut bapak Kapolri Jenderal Polisis Tito Karnavian menyampaikan bahwa dalam kurun 1 atau 2 hari berkas perkara beserta barang bukti tersangka Basuki Tjahaja Purnama “Ahok” akan diserahkan pada kejaksaan setempat agar dapat diproses sesuai ketentuan hukum yang berlaku. Pihak GNPFMUI di wakikan oleh ketua Pembina GNPF_MUI yaitu Rizieq Shihab mengutarakan bahwa aksi ini akan dilakukan secara super damai dimana massa melaksanakan aksi yang telah direncanakan dengan khitman dan damai.

METODE PENELITIAN

Penelitian ini bertujuan untuk menghasilkan akurasi yang tinggi dalam opini Kesepakatan Polri dan GNPF-MUI pada Aksi Bela Islam III '212' menggunakan algoritma Naïve Bayes sebab memiliki kelebihan ialah simpel namun menciptakan akurasi yang besar walaupun memakai informasi yang sedikit. Kelemahan dari naive bayes yakni hasil dari probabilitas kadangkala tidak normal serta masih kurang nya fitur yang cocok terhadap klasifikasi sehingga membuat akurasi sedikit rendah.

HASIL DAN PEMBAHASAN

Data mining

Text mining adalah suatu proses di mana komputer digunakan untuk menemukan informasi baru atau mengidentifikasi kembali informasi yang tersirat secara implisit dari sumber data teks yang berbeda. Salah satu bagian dari text mining adalah text preprocessing, yang bertujuan untuk menghilangkan noise pada kalimat dan menghindari data yang kurang sempurna, gangguan pada data, dan data yang tidak konsisten. Text preprocessing sangat

penting dalam analisis sentimen, yang merupakan teknik pengolahan data teks secara otomatis untuk mengevaluasi dan mengidentifikasi sentimen atau opini yang terkandung dalam teks seperti artikel, review produk, tweet, atau postingan media sosial lainnya. Dalam analisis sentimen, teks diklasifikasikan ke dalam tiga kategori sentimen yaitu positif, negatif, atau netral. Teknik ini dapat membantu perusahaan untuk memahami opini publik terhadap produk atau layanan mereka, dan mengembangkan strategi pemasaran yang efektif. Ada beberapa metode yang dapat digunakan dalam analisis sentimen, seperti pendekatan berbasis aturan dan pendekatan berbasis mesin.

Ulasan Tokoh Masyarakat

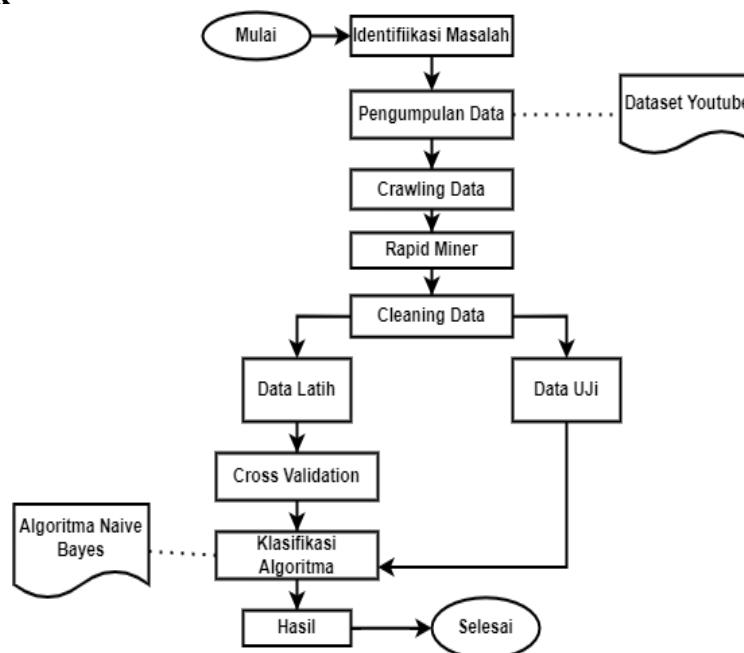
Aksi ini diawali sebagai bentuk protes masyarakat terhadap dugaan penistaan agama yang dilakukan gubernur Basuki Tjahaja Purnama “Ahok”. Untuk mempermudah dalam mengamati permasalahan ini media sosial Youtube digunakan untuk mengungkap opini masyarakat.

Naïve Bayes

Metode Naive Bayes adalah suatu teknik pengklasifikasian yang digunakan dalam machine learning. Metode ini didasarkan pada teori probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu dengan memprediksi probabilitas dimasa depan berdasarkan pengalaman dimasa sebelumnya. Konsep dasar teori Bayes adalah peluang bersyarat $P(X|C)$, di mana X adalah posterior dan C adalah prior. Prior adalah pengetahuan kita tentang karakteristik suatu parameter atau pengalaman dimasa lalu, sedangkan posterior adalah karakteristik yang diduga pada kejadian yang datang. Metode Naive Bayes menggunakan persamaan untuk menghitung probabilitas kelas yang paling mungkin berdasarkan fitur-fitur yang ada. Metode ini sering digunakan dalam analisis sentimen, spam filtering, recommendation systems, dan lainnya. Ada beberapa jenis Naive Bayes Classifier, seperti Multinomial Naive Bayes, Bernoulli Naive Bayes, dan Gaussian Naive Bayes, yang masing-masing cocok untuk jenis data yang berbeda.

$$p(C_j|d_i) = \frac{p(c_j|d_i)p(c_j)}{p(d_i)}$$

Tahapan Proyek



Gambar 1. Tahapan Proyek

Crawling Data

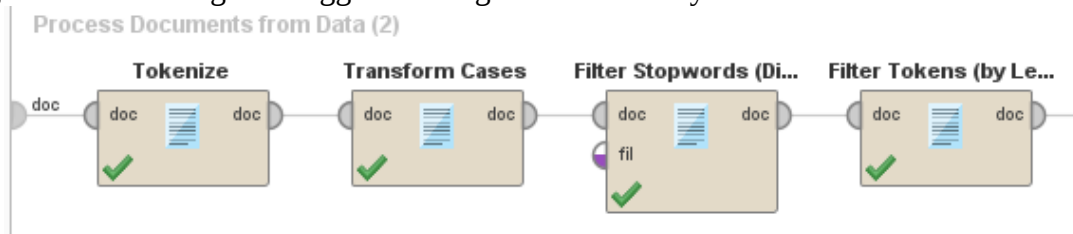
Pengumpulan informasi melalui penerapan teknologi crawler dapat secara otomatis mengumpulkan data. Crawler yang diperlukan harus disesuaikan dengan karakteristik khusus dari setiap media social [7]. pada Dataset telah dilakukan crawling menggunakan bantuan <https://netlytic.org/> dari hasil opini tersebut menghasilkan 873 komentar positif dan negative yang dilabeli secara manual.

text	sentimen
anehnya knp kl ada kejadian kyk gini para ormas dan lain lain gk ikut jd relawan untuk bela palestina	negatif
perbatasan itu seharusnya dengan sungai bkn tembok atau kawat sama sama aman	negatif
ya allahhh lindungilah saudara saudaraku kuatkan lah dan lindundungilah	positif
hilangkan judi online squad	positif
web judi masih meraja lela terutama squad gua udah rugi banyak emang awal menang	positif
kemarin saya merugi main judi di squad awal memang menang tapi kesininya saya rugi	negatif
semangat terus pak jangan lupa hapus judi online squad	negatif
berantas judi online squad masih aktif aja	positif
pesan saya tolong hapus judi online terutama squad	negatif
hapus judi online web udah lama masih aktif aja squad	positif
buar rugi banda judi online squad member baru di kasih jp terus	negatif
saya pernah main judi di squad awal saya menang tapi kesininya saya merugi tolong bantu pak	negatif
jangan tutup mata situs judi berkeliaran terutama dari squad	negatif
hmmm berantas judi online tuh web dari masih ada squad	positif
tolong berantas judi online terutama squad tuh masih aktif aja	positif

Gambar 2. Crawling Data

Pre-Processing

Preprocessing merupakan suatu langkah untuk mengubah istilah-istilah dalam kalimat menjadi bentuk yang normal. Tujuannya adalah agar data latih dan fitur-fitur yang diekstrak dapat disesuaikan secara konsisten dengan yang diinginkan, sehingga mempermudah proses pengolahan data. dengan menggunakan algoritma naïve bayes.



Gambar 3. Pre-Processing

TF-IDF

TF-IDF merupakan metrik statistik yang digunakan untuk menilai sejauh mana suatu kata penting dalam suatu dokumen. Komponen "TF" menggambarkan frekuensi kemunculan kata di dalam dokumen tertentu, mengindikasikan signifikansinya dalam dokumen tersebut. Sementara itu, "DF" mencerminkan frekuensi dokumen yang mengandung kata tersebut, memberikan gambaran tentang seberapa umum kata tersebut. "IDF" merupakan invers dari nilai DF. Hasil perhitungan TF-IDF diperoleh dengan mengalikan nilai TF dengan nilai IDF. Bobot kata akan semakin besar jika kata tersebut sering muncul dalam dokumen tertentu, dan sebaliknya, bobot akan semakin kecil jika kata tersebut muncul dalam banyak dokumen [9]. TF-IDF sebagai berikut:

aaamiin	aaamiinn	aalamiin	aamiin	aamiindan	aamiinn	abadi	abang
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0.150	0.261	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

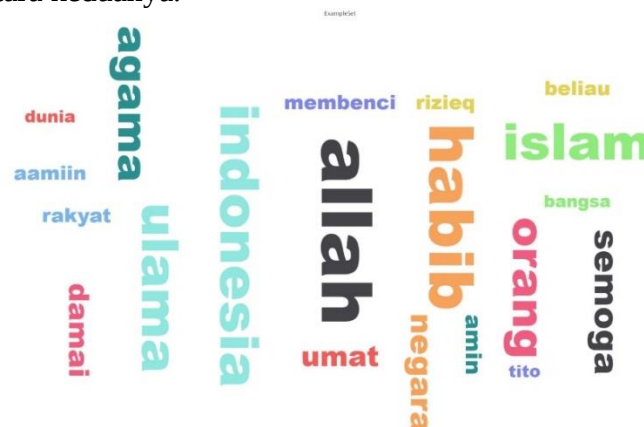
Gambar 5. Result TF-IDF

Row No.	word	in documents	total
1	allah	100	133
2	habib	92	119
3	indonesia	67	102
4	islam	76	101
5	ulama	72	99
6	orang	61	86
7	agama	52	80
8	semoga	53	69
9	negara	45	59
10	damai	47	55
11	umat	39	55
12	amin	37	46
13	membenci	10	42
14	rizieq	34	40
15	aamiin	36	38

Gambar 6. Result All TF-IDF

Word Cloud

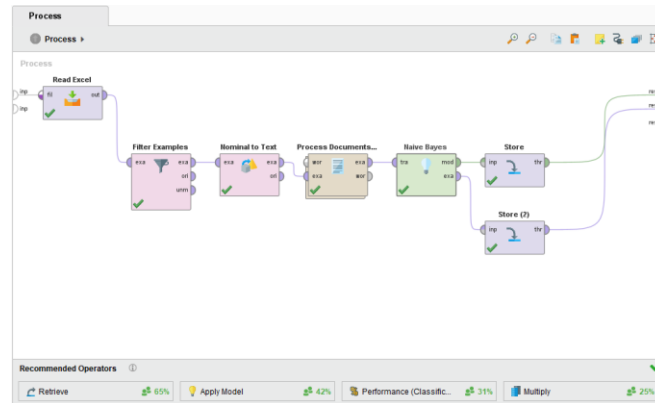
Word cloud adalah representasi visual yang intuitif dari daftar kata-kata yang digunakan dalam suatu teks. Secara umum, semakin sering sebuah kata muncul dalam teks, semakin besar representasinya dalam bentuk gambar. Dalam konteks ini, word cloud dapat berfungsi sebagai alat bantu visual untuk citraan, dan juga dikenal sebagai tag cloud atau text cloud. Ini adalah metode untuk menyorot bagian paling relevan dari data tekstual. Selain itu, word cloud dapat membantu dalam perbandingan dan perbedaan antara dua bagian teks untuk menemukan kesamaan kata di antara keduanya.



Gambar 7. Word Cloud

Naïve Bayes Classifier

Algoritma yang digunakan untuk analisis ini ialah naïve bayes. Digunakan untuk melihat akurasi ketika melakukan proses analisis sentiment, dengan menggunakan aplikasi yaitu Rapidminer maka akan menghasilkan.



Gambar 8. Process Naïve Bayes Classifier

Pada algoritma naïve bayes data set yang digunakan adalah 873 merupakan total dari proses pembersihan sebelumnya. Selanjutnya data dibagi menjadi data latih % dan data uji, %. Hasil nilai akurasi berada pada angka 89,70% seperti gambar di bawah:

Hasil Data:

accuracy: 89.70%			
	true positif	true negatif	class precision
pred. positif	394	0	100.00%
pred. negatif	62	146	70.19%
class recall	86.40%	100.00%	

precision: 70.19% (positive class: negatif)			
	true positif	true negatif	class precision
pred. positif	394	0	100.00%
pred. negatif	62	146	70.19%
class recall	86.40%	100.00%	

recall: 100.00% (positive class: negatif)			
	true positif	true negatif	class precision
pred. positif	394	0	100.00%
pred. negatif	62	146	70.19%
class recall	86.40%	100.00%	

Gambar 9. Hasil Data

KESIMPULAN

Hasil menunjukkan bahwa sistem atau model memiliki hasil yang baik. Dalam mengklasifikasi kelas positif dan negative memiliki accuracy 89,70%. Pada kelas positif hasil recall sebesar 86,40% artinya ada kasus positif yang tidak terdeteksi oleh sistem atau model. Untuk kelas negative precision sebesar 70,19% yang artinya hasil dari kelas negative mungkin salah. Namun recall yang ada pada kelas negative 100% artinya semua kasus kelas negative dapat terdeteksi dengan benar.

DAFTAR PUSTAKA

- M. R. Fanani, "ALGORITMA NAÏVE BAYES BERBASIS FORWARD SELECTION UNTUK PREDIKSI BIMBINGAN KONSELING SISWA Jurnal DISPROTEK," J. DISPROTEK, vol. 11, hal. 13–22, 2020.
- F. Kurniawan dan Q. Al Qorni, "Exploring Sentimen Analysis Using Machine Learning : A Case Study on Partai Demokrasi Indonesia Perjuangan (PDIP) in the 2024 General

- Election,” *sinomics juournal*, vol. 2, no. 4, hal. 911–920, 2023.
- F. Sidik, I. Suhada, A. H. Anwar, dan F. N. Hasan, “Analisis Sentimen Terhadap Pembelajaran Daring dengan Algoritma Naïve Bayes Classifier,” *J. Linguist. komputasional*, vol. 5, no. 1, hal. 34–43, 2022.
- V. K. S. Que, A. Iriani, dan H. D. Purnomo, “Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization (Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization),” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 9, no. 2, hal. 162–170, 2020.
- F. V. Sari dan A. Wibowo, “Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi,” *J. SIMETRIS*, vol. 10, no. 2, hal. 681–686, 2019.
- M. N. Muttaqin dan I. Kharisudin, “Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine dan K Nearest Neighbor,” *UNNES J. Math.*, vol. 10, no. 2, hal. 22–27, 2021, [Daring]. Tersedia pada: <http://journal.unnes.ac.id/sju/index.php/ujm>
- J. Budiarto, “Identifikasi Kebutuhan Masyarakat Nusa Tenggara Barat pada Pandemi Covid-19 di Media Sosial dengan Metode Crawling (Requirements Identification for NTB People in pandemic covid-19 at Social Media Using Crawling Method),” vol. 2, no. 4, hal. 244–250, 2021.
- O. I. Gifari, M. Adha, I. R. Hendrawan, F. Freddy, dan S. Durrand, “Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine,” *JIFOTECH (JOURNAL Inf. Technol.*, vol. 2, no. 1, hal. 36–40, 2022.
- J. A. Septian, T. M. Fahrudin, dan A. Nugroho, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF - IDF dan K - Nearest Neighbor,” *J. Intell. Syst. Comput.*, hal. 43–49, 2019.
- T. M. Fahrudin, A. Ruhui, F. Sari, A. Lisanthoni, dan A. A. Dewi, “Analisis Speech-to-Text pada Video Mengandung Kata Kasar dan Ujaran Kebencian dalam Ceramah Agama Islam Menggunakan Interpretasi Audiens dan Visualisasi Word Cloud,” *SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 5, hal. 190–202, 2022.