

Analisis Sentimen Publik Terhadap RUU KUHAP di Platform X Menggunakan Metode TF-IDF dan Naïve Bayes

Junaidy¹, Muhammad Fauzan², Roberto Kaban³

^{1,2}Jurusan Rekayasa Perangkat Lunak, Institut Teknologi dan Bisnis Indonesia

³Jurusan Teknik Informatika, Institut Teknologi dan Bisnis Indonesia

E-mail:¹nedyzulkarnain33@gmail.com, ²fauzanmohd336@gmail.com,
³robertokaban@itbi.ac.id

Intisari

Perkembangan media sosial menjadikan platform X sebagai salah satu sarana utama masyarakat dalam menyampaikan opini terhadap isu-isu kebijakan publik, termasuk Rancangan Undang-Undang Kitab Undang-Undang Hukum Acara Pidana (RUU KUHAP). Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap RUU KUHAP berdasarkan data tweet yang dikumpulkan dari platform X. Diperoleh 2.273 data valid yang digunakan dalam penelitian. Data terpilih kemudian melalui tahapan preprocessing yang meliputi case folding, cleansing, tokenizing, stopword removal, dan stemming. Ekstraksi fitur dilakukan menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF), sedangkan proses klasifikasi sentimen dilakukan menggunakan algoritma Multinomial Naive Bayes dengan pembagian data menjadi data latih dan data uji. Evaluasi performa model dilakukan menggunakan confusion matrix serta metrik precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa sentimen publik terhadap RUU KUHAP didominasi oleh sentimen negatif. Hasil analisis sentimen menunjukkan bahwa opini publik terhadap RUU KUHAP didominasi oleh sentimen negatif sebesar 45,5%, diikuti oleh sentimen netral sebesar 32,0%, dan sentimen positif sebesar 22,5%. Evaluasi performa model menunjukkan bahwa pada kelas negatif diperoleh nilai precision 0,71, recall 0,93, dan F1-score 0,80, pada kelas netral precision 0,74, recall 0,44, dan F1-score 0,55, serta pada kelas positif precision 0,85, recall 0,80, dan F1-score 0,82. Secara keseluruhan, model mencapai tingkat akurasi sebesar 74.07%, yang menunjukkan bahwa penerapan TF-IDF dan Naïve Bayes mampu mengklasifikasikan sentimen publik dengan cukup baik meskipun masih terdapat keterbatasan dalam mengidentifikasi sentimen netral.

Kata Kunci : Analisis Sentimen, RUU KUHAP, Platform X (Twitter), Text Mining, TF-IDF, Opini Publik

**Analisis Sentimen Publik Terhadap RUU KUHAP di Platform X Menggunakan Metode
TF-IDF dan Naïve Bayes
(Junaidy, Muhammad Fauzan, Roberto Kaban)**

Abstract

The rapid development of social media has established Platform X as one of the primary channels for the public to express opinions on public policy issues, including the Draft Criminal Procedure Code (RUU KUHAP). This study aims to analyze public sentiment toward the RUU KUHAP based on tweet data collected from Platform X. A total of 2,273 valid data points were obtained and utilized in this research. The selected data underwent several preprocessing stages, including case folding, cleansing, tokenizing, stopword removal, and stemming. Feature extraction was performed using the Term Frequency–Inverse Document Frequency (TF-IDF) method, while the sentiment classification process employed the Multinomial Naive Bayes algorithm, with the dataset split into training and testing sets. Model performance was evaluated using a confusion matrix alongside precision, recall, and F1-score metrics. The results indicate that public sentiment toward the RUU KUHAP is dominated by negative sentiment at 45.5%, followed by neutral sentiment at 32.0%, and positive sentiment at 22.5%. Performance evaluation shows that for the negative class, the model achieved a precision of 0.71, recall of 0.93, and F1-score of 0.80. For the neutral class, the precision was 0.74, recall 0.44, and F1-score 0.55, while the positive class reached a precision of 0.85, recall 0.80, and F1-score 0.82. Overall, the model achieved an accuracy of 74.07%, demonstrating that the application of TF-IDF and Naive Bayes is effective in classifying public sentiment, despite persistent limitations in identifying neutral sentiment.

Keywords : Sentiment Analysis, RUU KUHAP, Platform X (Twitter), Text Mining, TF-IDF, Public Opinion

PENDAHULUAN

Kemajuan teknologi informasi dan komunikasi telah membawa perubahan mendasar dalam pola interaksi sosial masyarakat, terutama melalui pemanfaatan media sosial. Platform digital memungkinkan individu untuk berpartisipasi dalam diskursus publik, menyampaikan pandangan, serta merespons isu-isu sosial dan politik secara langsung dan berkelanjutan. Dalam konteks ini, Twitter yang kini dikenal sebagai X menempati posisi strategis sebagai ruang ekspresi opini publik karena karakteristiknya yang terbuka, cepat, dan mampu menampung keragaman perspektif masyarakat terhadap kebijakan pemerintah maupun agenda legislatif yang sedang berlangsung [1].

Seiring dengan meningkatnya volume data teks yang dihasilkan di media sosial, analisis sentimen berkembang sebagai salah satu pendekatan utama dalam ranah *text mining* dan *natural language processing* (NLP). Pendekatan ini berfokus pada pengelompokan opini subjektif ke dalam kategori sentimen tertentu, seperti positif, negatif, atau netral. Berbagai penelitian telah menunjukkan bahwa analisis sentimen efektif digunakan untuk mengevaluasi respons publik terhadap layanan, produk, serta kebijakan publik dalam beragam konteks media sosial [2], [3]. Dalam proses tersebut, *feature extraction* berbasis *term frequency-inverse document frequency* (TF-IDF) terbukti mampu merepresentasikan kepentingan relatif suatu istilah dalam korpus teks secara lebih informatif, sehingga mendukung kinerja model klasifikasi sentimen [2].

Di antara berbagai algoritma klasifikasi yang tersedia, Naïve Bayes kerap digunakan dalam analisis teks karena struktur modelnya yang sederhana, kebutuhan komputasi yang relatif rendah, serta performa yang stabil pada data teks berlabel. Sejumlah studi empiris menunjukkan bahwa integrasi TF-IDF dengan Naïve Bayes mampu menghasilkan pemetaan sentimen yang cukup akurat dalam menganalisis opini publik terhadap isu-isu kebijakan, termasuk pada sektor layanan perbankan dan program pemerintah [4], [5]. Temuan-temuan tersebut mengindikasikan bahwa kombinasi kedua metode ini masih relevan dan kompetitif untuk studi opini publik .

RUU KUHAP sendiri merupakan instrumen hukum yang memiliki dampak signifikan terhadap sistem peradilan pidana, perlindungan hak asasi manusia, serta legitimasi proses penegakan hukum di mata publik. Ketika pembahasan suatu rancangan undang-undang memperoleh sorotan luas di media sosial, pemahaman yang berbasis data mengenai sentimen masyarakat menjadi semakin penting. Informasi tersebut dapat berfungsi sebagai masukan empiris bagi para pembuat kebijakan untuk memahami dinamika persepsi publik, sekaligus sebagai dasar dalam merumuskan kebijakan yang lebih adaptif dan responsif terhadap aspirasi social [6].

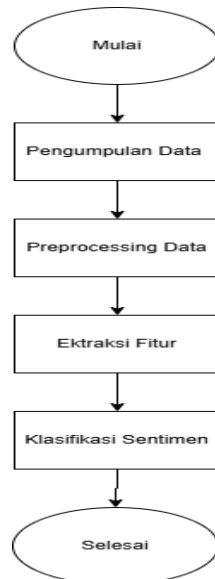
Berdasarkan pertimbangan tersebut, penelitian ini bertujuan untuk melakukan analisis sentimen publik terhadap Rancangan Undang-Undang Kitab Undang-Undang Hukum Acara Pidana (RUU KUHAP) dengan memanfaatkan data teks yang bersumber

Analisis Sentimen Publik Terhadap RUU KUHAP di Platform X Menggunakan Metode TF-IDF dan Naïve Bayes (Junaidy, Muhammad Fauzan, Roberto Kaban)

dari platform X. Analisis sentimen dilakukan dengan menerapkan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) sebagai teknik ekstraksi fitur dan algoritma *Naïve Bayes* sebagai metode klasifikasi. Hasil penelitian ini diharapkan dapat memberikan gambaran mengenai kecenderungan sentimen masyarakat terhadap RUU KUHAP serta menunjukkan efektivitas penerapan TF-IDF dan Naïve Bayes dalam analisis sentimen berbasis media sosial.

METODOLOGI PENELITIAN

Metodologi penelitian ini disusun secara sistematis untuk memastikan proses analisis sentimen berjalan terstruktur dan dapat dipertanggungjawabkan secara ilmiah. Metodologi diawali dengan tahap pengumpulan data dari platform X, yang menjadi sumber utama opini publik terkait RUU KUHAP. Data yang dikumpulkan berupa tweet yang relevan dengan topik penelitian berdasarkan kata kunci yang telah ditentukan. Tahap selanjutnya adalah preprocessing data, yang bertujuan untuk membersihkan dan menormalkan teks agar siap dianalisis. Tahap preprocessing data merupakan bagian penting dalam analisis sentimen karena data mentah, khususnya dari media sosial, sering mengandung noise seperti URL, mention, emotikon, dan karakter yang tidak relevan sesuai tujuan analisis. Preprocessing dilakukan untuk membersihkan dan menormalkan teks, sehingga dapat meningkatkan kualitas dan akurasi pemodelan [7], [8], [9].



Gambar 1. Alur Metode Penelitian

Tahapan penelitian dimulai dari pengumpulan data, dilanjutkan dengan preprocessing untuk membersihkan dan menormalkan teks agar siap dianalisis. Setelah data bersih, dilakukan ekstraksi fitur menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF memberikan bobot numerik pada setiap kata berdasarkan tingkat kepentingannya dalam dokumen dan seluruh dataset, sehingga representasi numerik ini dapat diproses oleh algoritma klasifikasi [10]. Tahap terakhir adalah klasifikasi menggunakan algoritma Multinomial Naive Bayes. Data dibagi menjadi data latih (*training*) dan data uji (*testing*). Model belajar dari pola kata yang sudah diberi label manual, kemudian memprediksi sentimen pada data uji untuk dievaluasi kinerjanya menggunakan *Confusion Matrix*.

1. Pengumpulan Data

Data penelitian diperoleh dari platform X dengan menggunakan teknik scraping data. Platform X dipilih karena merupakan salah satu media sosial yang paling aktif digunakan oleh masyarakat untuk menyampaikan opini terkait isu-isu hukum dan kebijakan publik [11], termasuk RUU KUHAP. Pengambilan data dilakukan dengan menentukan kata kunci yang relevan seperti "RUU KUHAP", "pengesahan KUHAP", dan "revisi KUHAP". Kata kunci tersebut digunakan untuk menyaring tweet agar data yang diperoleh sesuai dengan topik penelitian. Data yang telah dikumpulkan kemudian diseleksi kembali untuk memastikan relevansi dan menghilangkan komentar duplikat atau tidak relevan. Tahap ini penting agar kualitas data yang digunakan dalam analisis sentimen tetap terjaga [12].

2. Preprocessing Data

Tahapan *preprocessing* bertujuan untuk membersihkan dan menyiapkan data teks sebelum dilakukan analisis lebih lanjut. Data mentah dari platform X umumnya mengandung noise seperti URL, *mention*, *hashtag*, emoji, serta penggunaan bahasa informal yang tidak dibutuhkan dalam analisis model. Preprocessing merupakan langkah penting dalam analisis sentimen karena dapat mengurangi noise dan membuat data lebih terstruktur sehingga dapat meningkatkan performa model klasifikasi [13]. Langkah preprocessing yang dilakukan meliputi case folding untuk mengubah seluruh teks menjadi huruf kecil, cleansing untuk menghapus karakter yang tidak diperlukan, tokenizing untuk memecah teks menjadi kata-kata, stopword menghapus kata-kata umum yang memiliki frekuensi tinggi namun tidak mengandung muatan sentimen [14], [15]. Selain itu, proses stemming dilakukan untuk mengubah kata ke bentuk dasarnya. Dengan tahapan preprocessing ini, data teks menjadi lebih terstruktur dan siap digunakan dalam proses ekstraksi fitur, seperti berikut:

- a. Case Folding: Merupakan tahapan pengolahan teks yang bertujuan untuk menyamakan bentuk huruf dalam sebuah teks menjadi satu format yang konsisten, biasanya dengan mengubah semua huruf menjadi huruf kecil (lowercase), (misalnya: "RUU" dan "ruu") sebagai satu entitas yang identik [16].

**Analisis Sentimen Publik Terhadap RUU KUHAP di Platform X Menggunakan Metode
TF-IDF dan Naïve Bayes
(Junaidy, Muhammad Fauzan, Roberto Kaban)**

- b. **Cleansing:** merupakan tahapan membersihkan data dari elemen yang tidak relevan atau tidak berdampak pada proses klasifikasi pada data teks [16].
 - c. **Tokenizing:** Tahap ini memecah kalimat menjadi token per kata. Proses ini dilakukan dengan tujuan agar proses berikutnya dapat diterapkan pada tingkat kata [17].
 - d. **Stopword Removal:** Menghapus kata-kata umum yang sering muncul dalam kalimat namun tidak memberikan informasi penting terkait sentimen, seperti kata hubung ("dan", "di", "ke", "yang"). Proses ini membantu model berfokus pada kata kunci yang bermakna [17].
 - e. **Stemming:** merupakan tahapan untuk mengubah kata-kata modern atau singkatan menjadi bentuk dasar [18].
3. **Ekstraksi Fitur (TF-IDF)**

Ekstraksi fitur merupakan tahap untuk mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma klasifikasi. Metode yang digunakan dalam penelitian ini adalah TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF memberikan bobot pada setiap kata berdasarkan frekuensi kata dalam dokumen serta keberadaannya secara keseluruhan dalam korpus. Dengan metode ini, kata-kata yang sering muncul pada satu tweet tetapi jarang pada tweet lain akan memiliki bobot lebih tinggi, sehingga keterwakilannya dalam analisis sentimen lebih jelas [19].

4. **Klasifikasi Sentimen (Naïve Bayes)**

Tahap klasifikasi bertujuan untuk mengelompokkan tweet ke dalam kategori sentimen positif, negatif, atau netral. Proses klasifikasi dilakukan menggunakan algoritma Multinomial Naive Bayes yang terkenal efektif untuk text classification dengan data teks numerik berbasis TF-IDF [10], [20]. Model klasifikasi dilatih menggunakan data yang telah diberi label sentimen, kemudian digunakan untuk memprediksi sentimen pada data uji. Hasil klasifikasi ini menjadi dasar dalam analisis sentimen public terhadap RUU KUHAP. Evaluasi kinerja model dilakukan dengan confusion matrix, yang menampilkan jumlah prediksi benar dan salah per kelas, serta dihitung metrik seperti precision, recall, dan F1-score untuk menilai ketepatan model dalam mengklasifikasikan setiap kategori sentimen.

HASIL DAN PEMBAHASAN

1. Pengumpulan Data

Penelitian diawali dengan pengumpulan data dari media sosial X (Twitter) menggunakan kata kunci "RUU KUHP". Pada awalnya, berhasil dihimpun sebanyak 2.597 data mentah. Namun, untuk menjaga kualitas analisis, dilakukan proses penyaringan manual terhadap konten yang tidak relevan. Komentar yang bersifat iklan (*spam*), promosi, atau diskusi di luar konteks hukum RUU KUHP dieliminasi. Melalui tahap ini, sebanyak 324 entri dibuang, karena tidak berhubungan langsung dengan topik yang sedang dibahas, menyisakan 2.273 data valid yang kemudian diberi label sentimen (Positif, Negatif, Netral) secara manual sebagai acuan (*ground truth*) yang akan di proses tahap selanjutnya.

3	Mon Dec 22 07:40:03 +0000 2025	Bersama dukung RUU KUHP ciptakan sistem peradilan yang aman dan adil bagi semua rak in	https://x.com/undefined/status/2003007612116201491	positif
4	Mon Dec 22 05:42:10 +0000 2025	Bersama dukung RUU KUHP ciptakan sistem peradilan yang aman dan adil bagi semua rak in	https://x.com/undefined/status/2002977944441495800	positif
5	Mon Dec 22 06:10:30 +0000 2025	RUU KUHP persial hukum bagi setiap warga tanpa padang bulu https://t.co/Nu4qQL5W5g in	https://x.com/undefined/status/2002985078004253136	netral
6	Sun Dec 21 14:03:14 +0000 2025	RUU KUHP baru menekankan konsep restorative justice dan ruang pemaafan membuka g in	https://x.com/undefined/status/2002741657637041802	positif
7	Mon Dec 22 05:57:13 +0000 2025	RUU KUHP persial hukum bagi setiap warga tanpa padang bulu https://t.co/8iCC3k0E57 in	https://x.com/undefined/status/2002981731545280954	netral
8	Thu Nov 20 05:29:32 +0000 2025	DRP sahan RUU KUHP kemajuan kemunduran atau justru kehancuran ? Yg pasti ketidaka in	https://x.com/undefined/status/1991378355300758034	netral
9	Mon Dec 22 05:52:33 +0000 2025	RUU KUHP Persial Hukum Bagi Setiap Warga Tanpa Pandang Bulu https://t.co/OluMwACK in	https://x.com/undefined/status/200298055825721276	netral
10	Thu Dec 11 05:29:58 +0000 2025	RUU KUHP : biar hukum kita makin fair dan ngga ribet https://t.co/nHN4w4MIVC in	https://x.com/undefined/status/199898660933732828	netral
11	Sun Dec 21 15:27:59 +0000 2025	Bukan cuma soal hukum RUU KUHP juga bawa harapan baru buat disabilitas di Indonesia in	https://x.com/undefined/status/2002762982630650606	positif
12	Mon Dec 22 09:20:50 +0000 2025	Wujudkan sistem peradilan yang bersih dan berhaik nuansa. Dukung RUU KUHP sebagai p in	https://x.com/undefined/status/2002972579329949431	positif
13	Mon Nov 24 00:39:39 +0000 2025	Gimana mau dipercaya kalo selalu berseberangan dgn aspirasi rakyat? RUU TNI & KUH in	https://x.com/undefined/status/1992754554837243733	negatif
14	Sat Dec 06 02:22:24 +0000 2025	Peringatan Darurat Komisi III DPR RI dan Pemerintah telah menyetujui Rancangan Undang in	https://x.com/undefined/status/1997125468574564821	netral
15	Sat Dec 13 12:05:18 +0000 2025	Apa penadapatmu tentang pentingnya RUU KUHP untuk keadilan yang lebih pasti? Yuk dis in	https://x.com/undefined/status/1999812875141963799	netral
16	Sat Nov 22 11:40:00 +0000 2025	Retno Yudihi Bongkar Cara DPR Sembunyiin Pasai-Pasai Dalam Draft RUU KUHP https://t in	https://x.com/undefined/status/199119630507174050	netral
17	Sat Dec 06 04:43:12 +0000 2025	Reformasi Sistem Peradilan Pidana Indonesia melalui RUU KUHP demi peradilan yang tel in	https://x.com/undefined/status/1997130082046113115	positif
18	Thu Dec 04 11:15:59 +0000 2025	Dukung pengesahan RUU KUHP itu keadilan yang lebih manusawi https://t.co/ZL71m in	https://x.com/undefined/status/199653796340246491	positif
19	Wed Dec 03 09:07:52 +0000 2025	Inilah alasan mengapa RUU KUHP itu penting! https://t.co/40BlyWPHAD in	https://x.com/undefined/status/1996344344596918686	netral
20	Tue Dec 16 13:34:26 +0000 2025	RUU KUHP : hukum hadir untuk melindungi! https://t.co/2PqC5jg6g in	https://x.com/undefined/status/200092749969397729	positif
21	Mon Dec 22 05:55:49 +0000 2025	RUU KUHP Persial Hukum Bagi Setiap Warga Tanpa Pandang Bulu https://t.co/wk0w1WB81 in	https://x.com/undefined/status/200296087723226277	netral
22	Tue Dec 09 19:27:43 +0000 2025	Bersama Dukung RUU KUHP Wujudkan Hukum Pidana yang Modern Mari kita bersama-sa in	https://x.com/undefined/status/1998474661866978768	positif

Gambar

2. Data

Ground

Truth

2. Hasil Preprocessing Data

Tahap *preprocessing* merupakan fase krusial untuk mengubah teks tidak terstruktur menjadi format yang bersih dan bermakna. Alur pengolahan dilakukan sebagai berikut:

- Case Folding*: Seluruh teks diseragamkan menjadi huruf kecil (*lowercase*) agar tidak terjadi duplikasi fitur.

**Analisis Sentimen Publik Terhadap RUU KUHAP di Platform X Menggunakan Metode TF-IDF dan Naïve Bayes
(Junaidy, Muhammad Fauzan, Roberto Kaban)**

Tabel 1. Hasil Tahap *Case Folding*

No	Sebelum	Sesudah
1.	Klo ini benar ... Berarti ini adalah statement paling biadab. UU lain apakah kalian tgg rakyat setuju ? Soal KUHAP & KUHP apakah kalian menunggu rakyat setuju ? Knp RUU lain sangat mudah & cpt kalian sahkan sedang kan terkait perampasan aset sangat sulit ?	klo ini benar ... berarti ini adalah statement paling biadab. uu lain apakah kalian tgg rakyat setuju ? soal kuhap & kuhp apakah kalian menunggu rakyat setuju ? knp ruu lain sangat mudah & cpt kalian sahkan sedang kan terkait perampasan aset sangat sulit ?
2.	RUU KUHAP tidak izinkan penangkapan tanpa bukti. Draft RUU KUHAP tetap mensyaratkan minimal dua alat bukti serta persetujuan penyidik. Waspada Hoaks yang beredar!!	ruu kuhap tidak izinkan penangkapan tanpa bukti. draft ruu kuhap tetap mensyaratkan minimal dua alat bukti serta persetujuan penyidik. waspada hoaks yang beredar!!
....
2273.	Pasal Karet di KUHP dan KUHAP Terbaru KUHP dan KUHAP terbaru berisi pasal karet yang mengancam kebebasan berpendapat. Berpotensi menjadi alat baru untuk kriminalisasi masyarakat. #TukangKupasPerkara #Tempodotco	pasal karet di kuhp dan kuhap terbaru kuhp dan kuhap terbaru berisi pasal karet yang mengancam kebebasan berpendapat. berpotensi menjadi alat baru untuk kriminalisasi masyarakat. #tukangkupasperkara #tempodotco

- b. *Cleansing* Menghapus elemen pengganggu seperti URL, *mentions* (@user), hashtag, tanda baca, dan angka.

Tabel 2. Hasil Tahap *Cleansing*

No	Sebelum	Sesudah
1.	klo ini benar ... berarti ini adalah statement paling biadab.	klo ini benar berarti ini adalah statement paling biadab uu lain

	uu lain apakah kalian tgg rakyat setuju ? soal kuhap & kuhp apakah kalian menunggu rakyat setuju ? knp ruu lain sangat mudah & cpt kalian sahkan sedang kan terkait perampasan aset sangat sulit ?	apakah kalian tgg rakyat setuju soal kuhap kuhp apakah kalian menunggu rakyat setuju knp ruu lain sangat mudah cpt kalian sahkan sedang kan terkait perampasan aset sangat sulit
2.	ruu kuhap tidak izinkan penangkapan tanpa bukti. draft ruu kuhap tetap mensyaratkan minimal dua alat bukti serta persetujuan penyidik. waspada hoaks yang beredar!!	ruu kuhap tidak izinkan penangkapan tanpa bukti draft ruu kuhap tetap mensyaratkan minimal dua alat bukti serta persetujuan penyidik waspada hoaks yang beredar
....
2273.	pasal karet di kuhp dan kuhap terbaru kuhp dan kuhap terbaru berisi pasal karet yang mengancam kebebasan berpendapat. berpotensi menjadi alat baru untuk kriminalisasi masyarakat. #tukangkupasperkara #tempodotco	pasal karet di kuhp dan kuhap terbaru kuhp dan kuhap terbaru berisi pasal karet yang mengancam kebebasan berpendapat berpotensi menjadi alat baru untuk kriminalisasi masyarakat

- c. *Tokenizing* (Tahap Baru): Setelah teks dicleaning, dilakukan proses Tokenization untuk memecah kalimat menjadi satuan kata tunggal atau *tokens*.

Tabel 4. Hasil Tahap *Tokenizing*

No	Sebelum	Sesudah
1.	klo ini benar berarti ini adalah statement paling biadab uu lain apakah kalian ttg rakyat setuju soal kuhap kuhp apakah kalian menunggu rakyat setuju kenapa ruu lain sangat mudah cpt kalian sahkan sedang kan terkait perampasan aset sangat sulit	['klo', 'ini', 'benar', 'berarti', 'ini', 'adalah', 'statement', 'paling', 'biadab', 'uu', 'lain', 'apakah', 'kalian', 'ttg', 'rakyat', 'setuju', 'soal', 'kuhap', 'kuhp', 'apakah', 'kalian', 'menunggu', 'rakyat', 'setuju', 'kenapa', 'ruu', 'lain', 'sangat', 'mudah', 'cpt', 'kalian', 'sahkan', 'sedang', 'kan', 'terkait',

**Analisis Sentimen Publik Terhadap RUU KUHAP di Platform X Menggunakan Metode TF-IDF dan Naïve Bayes
(Junaidy, Muhammad Fauzan, Roberto Kaban)**

		'perampasan', 'aset', 'sangat', 'sulit']
2.	ruu kuhap tidak izinkan penangkapan tanpa bukti draft ruu kuhap tetap mensyaratkan minimal dua alat bukti serta persetujuan penyidik waspada hoaks yang beredar	['ruu', 'kuhap', 'tidak', 'izinkan', 'penangkapan', 'tanpa', 'bukti', 'draft', 'ruu', 'kuhap', 'tetap', 'mensyaratkan', 'minimal', 'dua', 'alat', 'bukti', 'serta', 'persetujuan', 'penyidik', 'waspada', 'hoaks', 'yang', 'beredar']
....
2273.	pasal karet di kuhp dan kuhap terbaru kuhp dan kuhap terbaru berisi pasal karet yang mengancam kebebasan berpendapat berpotensi menjadi alat baru untuk kriminalisasi masyarakat	['pasal', 'karet', 'di', 'kuhp', 'dan', 'kuhap', 'terbaru', 'kuhp', 'dan', 'kuhap', 'terbaru', 'berisi', 'pasal', 'karet', 'yang', 'mengancam', 'kebebasan', 'berpendapat', 'berpotensi', 'menjadi', 'alat', 'baru', 'untuk', 'kriminalisasi', 'masyarakat']

- d. *Stopword Removal*: Menghapus kata-kata umum yang memiliki frekuensi tinggi namun tidak mengandung muatan sentimen (seperti: "adalah", "dari", "ke").

Tabel 5. Hasil Tahap *Stopword Removal*

No	Sebelum	Sesudah
1.	klo ini benar berarti ini adalah statement paling biadab uu lain apakah kalian ttg rakyat setuju soal kuhap kuhp apakah kalian menunggu rakyat setuju kenapa ruu lain sangat mudah cpt kalian sahkan sedang kan terkait perampasan aset sangat sulit	klo benar berarti statement paling biadab uu lain apakah kalian ttg rakyat setuju soal kuhap kuhp apakah kalian menunggu rakyat setuju kenapa ruu lain mudah cpt kalian sahkan sedang kan terkait perampasan aset sulit
2.	ruu kuhap tidak izinkan penangkapan tanpa bukti draft ruu kuhap tetap mensyaratkan minimal dua alat bukti serta	ruu kuhap tidak izinkan penangkapan tanpa bukti draft ruu kuhap tetap mensyaratkan minimal dua alat bukti

	persetujuan penyidik waspada hoaks yang beredar	persetujuan penyidik waspada hoaks beredar
....
2273.	pasal karet di kuhp dan kuhap terbaru kuhp dan kuhap terbaru berisi pasal karet yang mengancam kebebasan berpendapat berpotensi menjadi alat baru untuk kriminalisasi masyarakat	pasal karet kuhp kuhap terbaru kuhp kuhap terbaru berisi pasal karet mengancam kebebasan berpendapat berpotensi menjadi alat baru kriminalisasi masyarakat

- e. *Stemming*: Mengubah setiap token menjadi kata dasar menggunakan algoritma Sastrawi. Kami menerapkan proteksi token pada istilah "KUHAP" agar maknanya tidak berubah atau rusak akibat proses penghilangan imbuhan.

Tabel 6. Hasil Tahap *Stemming*

No	Sebelum	Sesudah
1.	['klo', 'ini', 'benar', 'berarti', 'ini', 'adalah', 'statement', 'paling', 'biadab', 'uu', 'lain', 'apakah', 'kalian', 'ttg', 'rakyat', 'setuju', 'soal', 'kuhap', 'kuhp', 'apakah', 'kalian', 'menunggu', 'rakyat', 'setuju', 'kenapa', 'ruu', 'lain', 'sangat', 'mudah', 'cpt', 'kalian', 'sahkan', 'sedang', 'kan', 'terkait', 'perampasan', 'aset', 'sangat', 'sulit']	klo ini benar berarti ini adalah pernyataan paling biadab uu lain apakah kalian ttg rakyat setuju soal kuhap kuhp apakah kalian tunggu rakyat setuju kenapa ruu lain sangat mudah cpt kalian sahkan sedang kan terkait perampasan aset sangat sulit

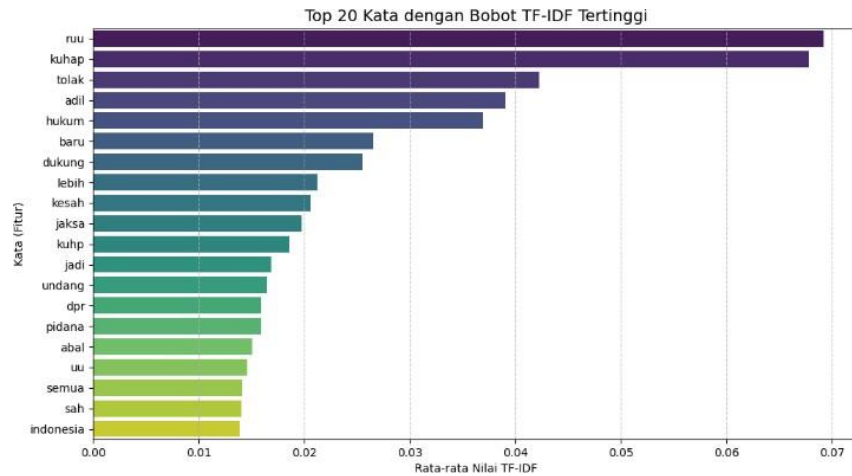
**Analisis Sentimen Publik Terhadap RUU KUHAP di Platform X Menggunakan Metode TF-IDF dan Naïve Bayes
(Junaidy, Muhammad Fauzan, Roberto Kaban)**

2.	['ruu', 'kuhap', 'tidak', 'izinkan', 'penangkapan', 'tanpa', 'bukti', 'draft', 'ruu', 'kuhap', 'tetap', 'mensyaratkan', 'minimal', 'dua', 'alat', 'bukti', 'serta', 'persetujuan', 'penyidik', 'waspada', 'hoaks', 'yang', 'beredar']	ruu kuhap tidak izin tangkap tanpa bukti draft ruu kuhap tetap syarat minimal dua alat bukti tuju sidik waspada hoaks edar
....
2273.	['pasal', 'karet', 'di', 'kuhp', 'dan', 'kuhap', 'terbaru', 'kuhp', 'dan', 'kuhap', 'terbaru', 'berisi', 'pasal', 'karet', 'yang', 'mengancam', 'kebebasan', 'berpendapat', 'berpotensi', 'menjadi', 'alat', 'baru', 'untuk', 'kriminalisasi', 'masyarakat']	pasal karet kuhp kuhap baru kuhp kuhap baru isi pasal karet ancam bebas dapat potensi jadi alat baru kriminalisasi masyarakat

3. Ekstraksi Fitur (TF-IDF)

Tahap krusial berikutnya adalah Ekstraksi Fitur menggunakan metode Pembobotan TF-IDF (*Term Frequency - Inverse Document Frequency*). Pada tahap ini, teks diubah menjadi nilai numerik yang merepresentasikan tingkat kepentingan suatu kata.

Berdasarkan visualisasi Top 20 Kata dengan Bobot TF-IDF Tertinggi, kata-kata inti yang mendominasi adalah "ruu" (0.069), "kuhap" (0.067), dan "tolak" (0.042). Tingginya bobot pada kata "tolak" mengindikasikan bahwa fitur penolakan merupakan komponen informasi yang sangat kuat dalam dataset ini.



Gambar 3. Top 20 Kata Dengan Bobot Tertinggi



Gambar 4. Word Cloud Berdasarkan Bobot

4. Hasil Klasifikasi Sentimen (Naive Bayes)

Model diklasifikasikan menggunakan algoritma Multinomial Naive Bayes dengan pembagian data 80% untuk pelatihan (*training*) dan 20% untuk pengujian (*testing*).

Tabel 7. Hasil *Classification Report*

	Precision	Recall	F1-Score	support
Negatif	0.71	0.93	0.80	210
Netral	0.74	0.44	0.55	152

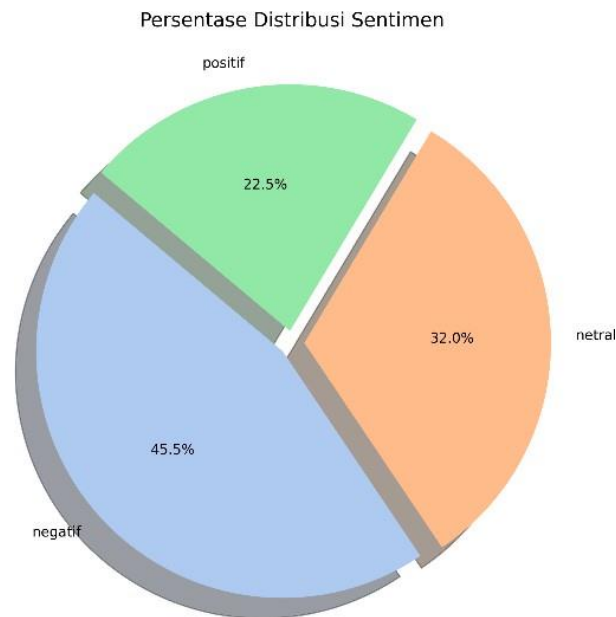
**Analisis Sentimen Publik Terhadap RUU KUHAP di Platform X Menggunakan Metode TF-IDF dan Naïve Bayes
(Junaidy, Muhammad Fauzan, Roberto Kaban)**

Positif	0.85	0.80	0.82	93
Accuracy			0.71	455

a. Distribusi Sentimen

Hasil pengolahan terhadap 2.273 data menunjukkan dominasi opini publik yang cenderung kritis terhadap RUU KUHAP:

- 1) Negatif (45,5%): Merupakan kategori terbesar dengan total 1.035 tweet.
- 2) Netral (32,0%): Terdiri dari 727 tweet yang umumnya bersifat informatif atau berupa berita.
- 3) Positif (22,5%): Terdiri dari 511 tweet yang berisi dukungan atau apresiasi.

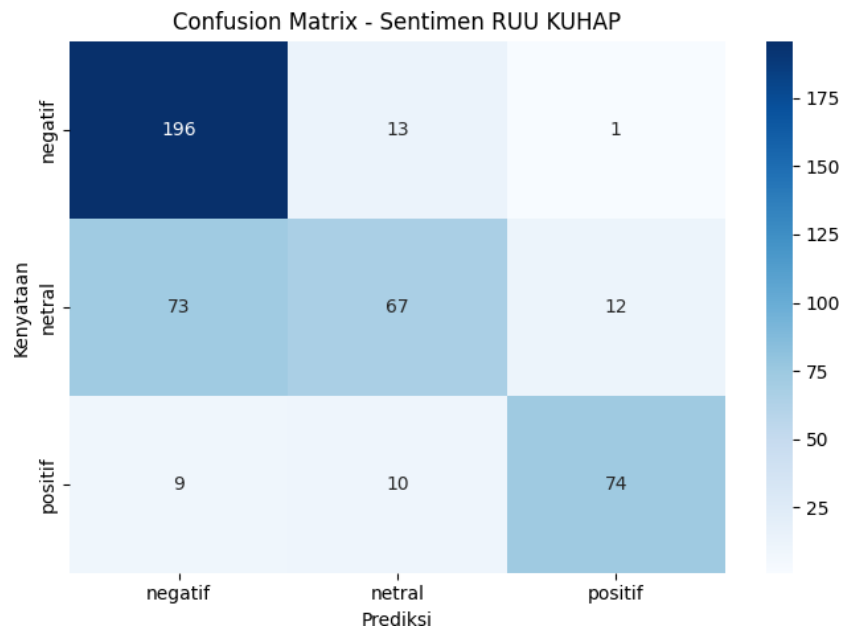


Gambar 5. Hasil Distribusi Sentimen

b. Evaluasi

Model Multinomial Naive Bayes diuji menggunakan *Confusion Matrix* untuk mengukur ketepatan prediksi pada data uji. Berdasarkan visualisasi yang diperoleh, model menunjukkan performa sebagai berikut:

- 1) Sentimen Negatif: Model berhasil memprediksi 196 data dengan benar (True Negative), dengan tingkat kesalahan yang sangat rendah.
- 2) Sentimen Netral: Sebanyak 67 data berhasil diprediksi benar, namun terdapat 73 data netral yang terdeteksi sebagai negatif oleh model.
- 3) Sentimen Positif: Sebanyak 74 data berhasil diprediksi secara akurat (True Positive).



Gambar 6. Convusion Matrix Sentimen

KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat ditarik beberapa kesimpulan sebagai berikut:

1. Opini Publik: Masyarakat di media sosial X menunjukkan kecenderungan respons yang kritis dan menolak terhadap RUU KUHAP. Hal ini dibuktikan dengan dominasi sentimen negatif sebesar 45,5% (1.035 tweet), diikuti oleh sentimen netral sebesar 32,0% (727 tweet), dan sentimen positif sebesar 22,5% (511 tweet).
2. Performa Model: Algoritma Multinomial Naive Bayes terbukti efektif dalam melakukan klasifikasi sentimen, terutama pada kategori negatif dengan hasil prediksi benar sebanyak 196 data pada data uji. Meskipun terdapat tantangan dalam membedakan antara kategori netral dan negatif, secara keseluruhan model memiliki kemampuan prediksi yang baik dengan dukungan pembobotan fitur yang akurat.

**Analisis Sentimen Publik Terhadap RUU KUHAP di Platform X Menggunakan Metode
TF-IDF dan Naïve Bayes
(Junaidy, Muhammad Fauzan, Roberto Kaban)**

SARAN

Adapun saran untuk menyempurnakan sistem yang telah dibuat adalah sebagai berikut:

1. Penelitian selanjutnya disarankan untuk menggunakan jumlah data yang lebih besar serta memperluas rentang waktu pengambilan data agar dapat menangkap dinamika opini publik yang lebih beragam terhadap isu kebijakan publik seperti RUU KUHAP. Dengan dataset yang lebih besar dan variatif, diharapkan performa model klasifikasi sentimen dapat ditingkatkan.
2. Penelitian selanjutnya dapat mempertimbangkan penggunaan metode atau algoritma klasifikasi lain, seperti Support Vector Machine (SVM), Random Forest, atau pendekatan deep learning, sebagai pembanding terhadap algoritma Naive Bayes. Perbandingan beberapa algoritma diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai metode terbaik dalam mengklasifikasikan sentimen publik.
3. Proses pelabelan data pada penelitian ini masih dilakukan secara manual. Oleh karena itu, penelitian selanjutnya dapat mengembangkan metode pelabelan semi-otomatis atau melibatkan lebih dari satu anotator untuk meningkatkan konsistensi dan objektivitas label sentimen.

DAFTAR PUSTAKA

- [1] W. B. Zulfikar, A. R. Atmadja, and S. F. Pratama, "Sentiment Analysis on Social Media Against Public Policy Using Multinomial Naive Bayes," *Scientific Journal of Informatics*, vol. 10, no. 1, pp. 25–34, 2023, doi: 10.15294/sji.v10i1.39952.
- [2] T. Aristianti, "Analisis sentimen kebijakan publik menggunakan algoritma Naive Bayes," pp. 1373–1384, 2020, doi: 10.33364/algoritma/v.22-2.2671.
- [3] D. Yuliani and R. Fadhila, "Sentiment Analysis of Public Opinion on National Debt Policy," *ICT Forum*, vol. 22, no. 2, 2023.
- [4] A. A. Hisyam and A. T. Ayunda, "BANK DKI PADA TWITTER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER," vol. 14, no. 1.
- [5] L. H. Sarumpaet and R. R. Suryono, "Analisis Sentimen Publik Program PPPK di Media Sosial X menggunakan Naïve Bayes dan SVM," *Edumatic: Jurnal Pendidikan Informatika*, vol. 9, no. 2, pp. 362–371, 2025, doi: 10.29408/edumatic.v9i2.30065.

- [6] F. Apri Wenando, “Analisis Opini Publik Terhadap Undang-Undang KUHP Tahun 2022 Menggunakan Algoritma Naïve Bayes Classifier,” *Jurnal Fasilkom*, vol. 13, no. 02, pp. 334–339, 2023, doi: 10.37859/jf.v13i02.5670.
- [7] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, “Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19),” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, p. 406, Apr. 2021, doi: 10.30865/mib.v5i2.2835.
- [8] E. Haddi, X. Liu, and Y. Shi, “The role of text pre-processing in sentiment analysis,” in *Procedia Computer Science*, Elsevier B.V., 2013, pp. 26–32. doi: 10.1016/j.procs.2013.05.005.
- [9] D. M. Putri and Y. Findawati, “Text Preprocessing on Emotional Tweets Case Study: Covid-19 Vaccine Rejection,” *Procedia of Engineering and Life Science*, vol. 7, pp. 446–453, Mar. 2024, doi: 10.21070/pels.v7i0.1501.
- [10] L. Hartimar, Y. Manza, and K. Putriani Siregar, “Text Classification Using TF-IDF and Naïve Bayes: Case Study of MyXL App User Review Data,” *Journal of Technology and Computer (JOTECHCOM)*, vol. 2, no. 2, pp. 100–108, 2025.
- [11] Y. Qi and Z. Shabrina, “Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach,” *Soc. Netw. Anal. Min.*, vol. 13, no. 1, Dec. 2023, doi: 10.1007/s13278-023-01030-x.
- [12] M. Faqih Rohmani, “Analisis Sentimen Pengguna Twitter Terhadap Universitas Pamulang Periode Penerimaan Mahasiswa Gelombang I Tahun Ajaran 2024/2025,” 2025.
- [13] N. Widya Utami, “TEXT MINING DALAM ANALISIS SENTIMEN PEMBELAJARAN DARING DI MASA PANDEMI COVID 19 MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR,” 2022.
- [14] L. Hermawan, M. B. Ismiati, J. Bangau, N. 60, and M. Charitas, “Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval,” *TRANSFORMATIKA*, vol. 17, no. 2, pp. 188–199, 2020.
- [15] G. Ainur, N. Latifah, and F. Nugraha, “Analisis Sentimen Publik Terhadap Program Makan Bergizi Gratis Menggunakan Metode Random Forest Pada Platform X Penulis Korespondensi.” [Online]. Available: <http://www.jurnal.umk.ac.id/sitech>
- [16] S. Andini, R. Kurniawan, and S. Anwar, “ANALISIS SENTIMEN PENGGUNA X MENGENAI OPINI MASYARAKAT TENTANG DINASTI POLITIK MENGGUNAKAN ALGORITMA NAÏVE BAYES,” *Jurnal*

**Analisis Sentimen Publik Terhadap RUU KUHAP di Platform X Menggunakan Metode
TF-IDF dan Naïve Bayes
(Junaidy, Muhammad Fauzan, Roberto Kaban)**

Informatika dan Teknik Elektro Terapan, vol. 13, no. 2, Apr. 2025, doi:
10.23960/jitet.v13i2.6299.

- [17] M. Alfonso and D. Bhisetya Rarasati, “JISA (Jurnal Informatika dan Sains) Sentiment Analysis of 2024 Presidential Candidates Election Using SVM Algorithm”.
- [18] Y. H. Agustin, N. Cici Mulyani, and W. Sindu Prasetya, “Analisis Sentimen Opini Publik Menggunakan Algoritma Naive Bayes dan TF-IDF,” *Jurnal Algoritma*, vol. 22, no. 2, Nov. 2025, doi: 10.33364/algoritma/v.22-2.2671.
- [19] U. N. Jakarta, V. B. Lestari, and C. A. Hutagalung, “J-KOMA Journal of Computer Science and Applications Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian-Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes,” pp. 22–2025, doi: 10.21009/j.
- [20] Cha Cha Kirana and Nabila Rizky Oktadini, “Analisis Sentimen Naïve Bayes dengan TF-IDF dan 10-Fold pada Ulasan Aplikasi X,” *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 7, no. 2, pp. 523–535, Dec. 2025, doi: 10.30865/json.v7i2.9007.