

ANALISIS SENTIMEN PUBLIK PADA TWITTER TERHADAP IMPLEMENTASI KEBIJAKAN ROYALTI MUSIK DENGAN ALGORITMA SUPPORT VECTOR MACHINE

Onky Ardhillah, Ari Wedhasmara, Ken Ditha Tania, Allsela Meiriza

Fakultas Ilmu Komputer, Universitas Sriwijaya, Jl. Raya Palembang - Prabumulih KM. 32, Indralaya Indah, Kec. Indralaya, Kabupaten Ogan Ilir, Sumatera Selatan 30862
onkyardhillah@gmail.com

ABSTRAK

Pemberlakuan Peraturan Pemerintah Nomor 56 Tahun 2021 tentang Pengelolaan Royalti Lagu dan/atau Musik bertujuan untuk melindungi hak ekonomi pencipta serta menciptakan sistem distribusi royalti yang adil. Namun, implementasi kebijakan ini menimbulkan polemik dan beragam respons dari masyarakat, khususnya pelaku usaha dan Usaha Mikro, Kecil, dan Menengah (UMKM), yang menilai kebijakan tersebut berpotensi menambah beban biaya serta masih kurang transparan dalam mekanisme penetapan tarif royalti. Kondisi ini menunjukkan adanya kesenjangan antara tujuan kebijakan dengan persepsi publik terhadap pelaksanaannya. Penelitian ini bertujuan untuk menganalisis dan mengklasifikasikan sentimen publik terhadap penerapan PP No. 56 Tahun 2021 melalui media sosial Twitter menggunakan pendekatan *text mining* dengan algoritma *Support Vector Machine* (SVM). Data penelitian diperoleh melalui proses *crawling* Twitter berdasarkan kata kunci terkait royalti dan musik pada periode Maret hingga November 2025 dengan total 3.813 tweet berbahasa Indonesia. Tahapan penelitian meliputi prapemrosesan teks yang terdiri dari *cleaning*, *case folding*, *normalization*, *tokenization*, *stopword removal*, dan *stemming*, kemudian dilakukan pelabelan sentimen ke dalam kelas positif, negatif, dan netral. Representasi fitur menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF), sedangkan proses klasifikasi dilakukan menggunakan algoritma SVM dengan kernel linier. Evaluasi model dilakukan menggunakan *confusion matrix* dengan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model SVM mampu mengklasifikasikan sentimen publik dengan tingkat akurasi sebesar 74,44%, dengan dominasi sentimen netral dan negatif dibandingkan sentimen positif. Temuan ini mengindikasikan bahwa sebagian masyarakat masih memiliki pandangan kritis terhadap implementasi kebijakan royalti musik, khususnya terkait transparansi dan mekanisme pengelolaannya, sehingga diperlukan peningkatan sosialisasi dan kejelasan tata kelola royalti oleh pemerintah dan Lembaga Manajemen Kolektif Nasional (LMKN).

Kata kunci : Analisis Sentimen, Text Mining, Support Vector Machine, Twitter, Royalti Musik.

1. PENDAHULUAN

Penerapan Peraturan Pemerintah Nomor 56 Tahun 2021 tentang Pengelolaan Royalti Lagu dan/atau Musik diarahkan untuk menjamin perlindungan hak ekonomi pencipta serta menciptakan mekanisme distribusi royalti yang berkeadilan [1]. Regulasi ini menetapkan kewajiban pembayaran royalti atas setiap penggunaan lagu atau musik dalam ruang publik kepada pencipta melalui sistem yang mengacu pada praktik internasional di industri musik [2]. Walaupun memiliki landasan tujuan yang kuat, implementasi PP No. 56 Tahun 2021 memicu kontroversi, terutama di kalangan pelaku usaha kecil yang memandang kewajiban royalti sebagai tambahan biaya yang tidak sebanding dengan kapasitas usaha mereka [3]. Keberatan tersebut tercermin dari pernyataan pelaku usaha kafe dan restoran yang menilai skema tarif royalti berdasarkan jumlah kursi berpotensi mengancam keberlanjutan usaha [4]. Dari sudut pandang Usaha Mikro, Kecil, dan Menengah (UMKM), ketentuan tarif royalti beserta aturan pendukungnya juga dianggap belum cukup terbuka, sehingga menyulitkan pelaku usaha dalam menentukan besaran kewajiban biaya secara pasti [3].

Situasi tersebut mengindikasikan adanya kesenjangan antara tujuan regulasi untuk melindungi

hak pencipta dengan persepsi masyarakat yang menilai kebijakan royalti masih membebani, kurang transparan, dan belum sepenuhnya mencerminkan keadilan [5]. Bagi UMKM, ketidakjelasan besaran tarif serta kompleksitas kerangka regulasi menjadi kendala utama dalam memperkirakan biaya yang harus dikeluarkan secara akurat [3]. Kesenjangan antara orientasi kebijakan dan respons publik ini perlu dikaji secara mendalam karena berpotensi memengaruhi tingkat penerimaan masyarakat yang pada akhirnya menentukan efektivitas pelaksanaan kebijakan publik [6]. Salah satu sumber penting untuk memahami respons tersebut adalah opini publik yang berkembang di media sosial, yang mencerminkan persepsi masyarakat secara langsung terhadap kebijakan yang diterapkan.

Permasalahan dalam penelitian ini tidak hanya berkaitan dengan polemik kebijakan royalti, tetapi juga mencakup tantangan dalam mengelola opini publik yang tersebar luas di media sosial. Opini publik tersebut umumnya berbentuk data teks tidak terstruktur, berjumlah besar, serta menggunakan bahasa informal dan singkatan, sehingga sulit dianalisis melalui pendekatan manual [7]. Penelitian terdahulu menunjukkan keterbatasan cakupan data, seperti studi yang dilakukan oleh Khawarga dan Riti

yang hanya menganalisis 300 tweet terkait polemik royalti Ahmad Dhani dan Once Mekel, sehingga belum mampu menggambarkan persepsi publik secara komprehensif. Kondisi ini menuntut penggunaan pendekatan *text mining* untuk melakukan proses ekstraksi, pembersihan, serta transformasi data teks menjadi informasi yang dapat dianalisis secara sistematis [8].

Di sisi lain, analisis sentimen berbasis teks juga menghadapi hambatan teknis, antara lain ketidakseimbangan distribusi kelas sentimen, seperti dominasi sentimen negatif, serta keterbatasan algoritma konvensional dalam mengolah data teks berskala besar [9]. Oleh sebab itu, pemilihan algoritma klasifikasi yang memiliki kinerja tinggi menjadi aspek krusial. *Support Vector Machine (SVM)* digunakan dalam penelitian ini karena kemampuannya dalam menangani data berdimensi tinggi dan performanya yang terbukti efektif dalam analisis sentimen berbasis teks [10]. Dengan demikian, fokus penelitian ini diarahkan pada pemanfaatan teknik *text mining* dan *machine learning* menggunakan algoritma SVM untuk memetakan persepsi publik terhadap kebijakan royalti musik secara objektif dan terukur [11]. Hasil penelitian ini menerapkan analisis sentimen berbasis media sosial Twitter dengan cakupan data yang lebih luas dan bersifat real-time untuk mengkaji persepsi publik terhadap PP No. 56 Tahun 2021. Hasil penelitian ini diharapkan dapat memberikan kontribusi akademik dalam pengembangan penerapan *text mining* dan *machine learning* berbasis SVM, sekaligus menjadi rujukan berbasis data bagi pemerintah dan Lembaga Manajemen Kolektif Nasional (LMKN) dalam memperbaiki tata kelola royalti musik di Indonesia.

2. TINJAUAN PUSTAKA

Untuk mendukung penelitian ini, diperlukan telaah terhadap berbagai penelitian dan sumber informasi lain yang membahas topik serupa, sehingga penelitian yang dilakukan dapat disusun berdasarkan landasan konseptual yang relevan, sistematis, dan selaras dengan permasalahan penelitian yang dikaji.

2.1. Penelitian Terdahulu

Kajian terdahulu menunjukkan bahwa analisis sentimen berbasis media sosial, khususnya Twitter, telah banyak digunakan untuk menilai respons publik terhadap berbagai isu sosial dan kebijakan dengan memanfaatkan algoritma machine learning. Khawarga dan Riti meneliti sentimen masyarakat terkait polemik royalti antara Ahmad Dhani dan Once Mekel menggunakan metode *Naïve Bayes* dengan jumlah data sebanyak 300 tweet. Penelitian tersebut menghasilkan tingkat akurasi sebesar 65,3%, yang dipengaruhi oleh keterbatasan jumlah data serta ambiguitas pada kelas sentimen netral [12].

Penerapan algoritma Support Vector Machine (SVM) pada penelitian lain menunjukkan performa yang lebih optimal ketika jumlah data meningkat. Safitri menggunakan 1.000 tweet terkait BTS dan

mencapai akurasi hingga 90%, meskipun analisis hanya mencakup dua kelas sentimen dan menghadapi permasalahan ketidakseimbangan data [13]. Temuan serupa disampaikan oleh Styawati yang menganalisis sentimen publik terhadap Program Kartu Prakerja dengan 2.000 tweet dan memperoleh akurasi di atas 98% menggunakan kernel linear, namun penelitian ini terbatas pada satu topik kebijakan dan periode pengamatan yang singkat [14].

Keterbatasan ruang lingkup data juga terlihat pada penelitian Ramlan serta Rifa yang masing-masing mengkaji isu kenaikan harga BBM dan layanan e-wallet menggunakan data Twitter dengan jumlah dan rentang waktu yang terbatas [15] [16]. Meskipun model SVM pada kedua penelitian tersebut menunjukkan tingkat akurasi yang relatif tinggi, keterbatasan variasi data dan fokus topik membatasi kemampuan generalisasi hasil.

Secara keseluruhan, penelitian-penelitian terdahulu mengindikasikan bahwa peningkatan jumlah data berpengaruh positif terhadap performa klasifikasi sentimen. Namun demikian, keterbatasan dataset, ketidakseimbangan distribusi kelas, serta sempitnya cakupan topik dan waktu masih menjadi kendala utama. Berdasarkan kondisi tersebut, penelitian ini diarahkan untuk melengkapi kekurangan penelitian sebelumnya dengan memanfaatkan data Twitter yang lebih luas dan bersifat *real-time* dalam menganalisis persepsi publik terhadap PP No. 56 Tahun 2021, dengan mengandalkan algoritma SVM yang terbukti efektif dalam pengolahan data teks berskala besar.

2.2. Hak Cipta Dan Royalti Musik

Hak cipta memberikan hak eksklusif kepada pencipta untuk mengatur pemanfaatan karya, termasuk penggandaan, pendistribusian, pertunjukan, dan pengumuman kepada publik [17]. Hak tersebut mencakup hak ekonomi yang memungkinkan pencipta memperoleh manfaat finansial, serta hak moral yang menjamin pengakuan dan perlindungan terhadap integritas karya [18]. Dalam kerangka hukum kekayaan intelektual, royalti dipahami sebagai bentuk imbalan finansial yang dibayarkan kepada pemegang hak atas penggunaan karya sesuai ketentuan peraturan perundang-undangan [19].

Pembayaran royalti atas pemanfaatan musik secara komersial maupun publik bertujuan memastikan terpenuhinya hak ekonomi pencipta, dengan pengelolaan dan distribusi yang dilakukan oleh LMKN dan LMK berdasarkan sistem data terintegrasi [20]. Meskipun demikian, penerapan sistem royalti dalam praktik sering menimbulkan perdebatan, khususnya terkait transparansi dan tingkat kepatuhan pelaku usaha kecil dan UMKM [21].

2.3. Regulasi Hak Cipta

Pemanfaatan lagu atau musik untuk kepentingan komersial dan publik di Indonesia mensyaratkan adanya izin dari pemegang hak cipta serta pemenuhan kewajiban pembayaran royalti sebagaimana diatur

dalam peraturan perundang-undangan di bidang kekayaan intelektual [22]. Regulasi tersebut menekankan pentingnya keberadaan repositori data lagu yang terpusat, lembaga resmi pengelola royalti, serta penerapan prinsip keadilan dan transparansi dalam penetapan tarif remunerasi [23]. Berbagai persoalan implementasi, seperti kejelasan prosedur, transparansi tarif, dan dampaknya terhadap UMKM, menjadikan pemetaan opini publik sebagai aspek penting dalam mengevaluasi efektivitas kebijakan [4, 5].

2.4. Twitter / X

Twitter (X) merupakan platform microblogging yang memungkinkan pengguna menyampaikan opini secara cepat dan ringkas, sehingga menjadi sumber data yang relevan untuk mengkaji persepsi publik terhadap isu kebijakan [24, 25]. Keunggulan utama Twitter terletak pada volume data yang besar, sifatnya yang aktual, serta kemudahan penelusuran berbasis kata kunci atau hashtag [26]. Namun, pemanfaatan data Twitter juga menghadapi sejumlah keterbatasan, seperti bias pengguna, keberadaan akun bot, dan noise linguistik, sehingga penelitian perlu memperhatikan aspek etika dan privasi dengan hanya menggunakan data publik serta menerapkan anonimisasi [27].

2.5. Analisis Sentimen

Analisis sentimen merupakan metode komputasional yang bertujuan untuk mengidentifikasi dan mengklasifikasikan kecenderungan sikap atau opini yang terkandung dalam data berbentuk teks ke dalam kategori sentimen tertentu, seperti positif, negatif, atau netral. Dalam penerapannya pada Bahasa Indonesia, analisis sentimen menghadapi tantangan linguistik berupa penggunaan bahasa informal, slang, serta ekspresi sarkastik yang kerap muncul dalam media sosial. Kondisi tersebut menyebabkan pendekatan *supervised learning* dengan representasi fitur yang tepat, seperti *Term Frequency-Inverse Document Frequency (TF-IDF)*, dinilai lebih efektif untuk menganalisis data tweet yang relatif singkat dan memiliki variasi bahasa yang tinggi

2.6. Text Mining

Text mining merupakan teknik pengolahan data teks yang berfokus pada ekstraksi dan transformasi data tidak terstruktur menjadi informasi yang bermakna dan dapat dianalisis secara sistematis. Dalam konteks analisis sentimen, *text mining* berperan sebagai fondasi utama karena mencakup serangkaian proses seperti pembersihan teks, normalisasi, serta representasi fitur yang memungkinkan teks mentah diolah lebih lanjut oleh algoritma klasifikasi. Melalui pendekatan ini, data tekstual dalam jumlah besar, khususnya dari media sosial, dapat dikonversi menjadi wawasan yang relevan untuk mendukung analisis opini publik.

2.7. Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan algoritma klasifikasi terawasi yang bekerja dengan membangun sebuah *hyperplane* optimal sebagai pemisah antar kelas data, dengan tujuan memaksimalkan margin antara support vector dari masing-masing kelas. Keunggulan SVM terletak pada fleksibilitas penggunaan fungsi kernel yang memungkinkan pemetaan data ke ruang berdimensi lebih tinggi, sehingga efektif dalam menangani data tekstual yang bersifat kompleks dan non-linear. Berbagai studi empiris menunjukkan bahwa SVM memiliki kinerja yang unggul dibandingkan algoritma klasifikasi lain dalam analisis sentimen berbasis Bahasa Indonesia, baik dari segi akurasi maupun stabilitas model.

3. METODE PENELITIAN

3.1. Pengambilan Data (Crawling)

Data penelitian dikumpulkan dari media sosial Twitter (X) melalui proses *crawling* menggunakan perangkat *Tweet Harvest* yang terhubung dengan API Twitter. Pengambilan data dilakukan secara otomatis berdasarkan kata kunci “royalti” dan “musik” dalam rentang waktu Maret hingga November 2025 untuk merepresentasikan diskursus publik terkait penerapan Peraturan Pemerintah Nomor 56 Tahun 2021 tentang Pengelolaan Royalti Lagu dan/atau Musik. Dataset dibatasi sebanyak 3.813 tweet guna menjaga keterwakilan opini publik dan efisiensi analisis, serta disimpan dalam format CSV setelah melalui pembersihan awal untuk menghilangkan elemen non-tekstual sebelum tahap pengolahan lanjutan.

3.2. Pembersihan Data (Cleaning)

Tahap *cleaning* data dilakukan untuk menghilangkan elemen non-tekstual dan *noise* pada data tweet, seperti URL, mention, hashtag, angka, emoji, dan tanda baca yang tidak relevan. Proses ini bertujuan meningkatkan kualitas data dengan memastikan bahwa teks yang dianalisis hanya mengandung informasi linguistik yang berkontribusi terhadap penentuan sentimen, sehingga mendukung akurasi tahap prapemrosesan dan klasifikasi selanjutnya.

3.3. Casefolding

Tahap *case folding* dilakukan dengan mengonversi seluruh huruf pada teks tweet ke dalam bentuk huruf kecil. Proses ini bertujuan menghilangkan perbedaan representasi kata akibat penggunaan huruf kapital, sehingga setiap kata diperlakukan secara konsisten dan tidak memengaruhi pembentukan fitur pada tahap analisis selanjutnya.

3.4. Normalization

Tahap *normalization* dilakukan untuk menyeragamkan penulisan kata pada teks tweet dengan mengubah ejaan tidak baku, singkatan, dan variasi bahasa informal ke dalam bentuk baku. Proses

ini bertujuan mengurangi variasi kata yang tidak relevan, sehingga meningkatkan konsistensi data dan kualitas representasi fitur pada tahap analisis selanjutnya.

3.5. Tokenization

Tahap *tokenization* dilakukan dengan memecah teks tweet menjadi unit-unit kata (token) berdasarkan spasi atau batas tertentu. Proses ini bertujuan mengubah teks yang semula berbentuk kalimat menjadi kumpulan kata yang dapat diproses secara komputasional, sehingga memudahkan tahap selanjutnya seperti penghapusan *stopword*, *stemming*, dan pembobotan fitur dalam analisis sentimen.

3.6. Stopword Removal

Tahap *stopword removal* dilakukan untuk menghilangkan kata-kata umum yang memiliki frekuensi tinggi namun tidak memberikan kontribusi signifikan terhadap penentuan sentimen. Penghapusan *stopword* bertujuan mengurangi dimensi data dan meningkatkan kualitas representasi teks dengan memfokuskan analisis pada kata-kata yang memiliki makna informatif.

3.7. Stemming

Tahap *stemming* dilakukan untuk mengubah setiap kata ke dalam bentuk dasarnya dengan menghilangkan imbuhan seperti awalan, sisipan, dan akhiran sesuai kaidah morfologi Bahasa Indonesia. Proses ini bertujuan menyatukan variasi kata yang memiliki makna serupa, sehingga meningkatkan konsistensi data teks dan kualitas representasi fitur sebelum tahap pembobotan dan klasifikasi sentimen.

3.8. Labeling

Tahap *labeling* dilakukan untuk mengelompokkan data tweet ke dalam kelas sentimen, yaitu positif, negatif, dan netral. Pelabelan ini digunakan sebagai dasar pembentukan data latih dalam proses klasifikasi, sehingga model dapat mempelajari pola karakteristik setiap kelas sentimen sebelum dilakukan pelatihan dan evaluasi model.

3.9. TF-IDF

Metode *Term Frequency-Inverse Document Frequency (TF-IDF)* digunakan untuk merepresentasikan data teks ke dalam bentuk numerik dengan memberikan bobot pada setiap kata berdasarkan tingkat kepentingannya dalam dokumen dan keseluruhan korpus. Kata yang sering muncul pada dokumen tertentu namun jarang muncul pada dokumen lain akan memperoleh bobot yang lebih tinggi, sedangkan kata yang muncul secara umum pada banyak dokumen akan memiliki bobot yang lebih rendah. Representasi ini memungkinkan teks diproses secara komputasional dan digunakan sebagai masukan pada model klasifikasi sentimen.

3.10. Support Vector Machine

Metode *Term Frequency-Inverse Document Frequency (TF-IDF)* digunakan untuk merepresentasikan data teks ke dalam bentuk numerik dengan memberikan bobot pada setiap kata berdasarkan tingkat kepentingannya dalam dokumen dan keseluruhan korpus. Kata yang sering muncul pada dokumen tertentu namun jarang muncul pada dokumen lain akan memperoleh bobot yang lebih tinggi, sedangkan kata yang muncul secara umum pada banyak dokumen akan memiliki bobot yang lebih rendah. Representasi ini memungkinkan teks diproses secara komputasional dan digunakan sebagai masukan pada model klasifikasi sentimen.

3.11. Confusion Matrix

Confusion matrix digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan hasil prediksi dan label aktual pada setiap kelas sentimen (positif, negatif, dan netral). Matriks ini menjadi dasar perhitungan metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score* untuk menilai keandalan model secara keseluruhan.

3.12. Wordcloud

Wordcloud digunakan sebagai teknik visualisasi untuk menampilkan kata-kata yang paling sering muncul dalam dataset setelah melalui tahap prapemrosesan. Ukuran kata pada wordcloud merepresentasikan frekuensi kemunculannya, sehingga membantu mengidentifikasi tema dominan dan kecenderungan umum opini publik dalam data teks.

4. HASIL DAN PEMBAHASAN

4.1. Pengambilan Data (Crawling)

	A	B	C	D	E	F	G	H	I	J	K	L
1	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_screen_name	lang	location			
2	19422362260274415	"Mon Jul 07 14:06:11 +0000 2025"	0	"NFT musik mulai jadi aset royalti digital. Artis bisa bagi hasil"	19422362260274415			"id"				
3	1942213471350976934	"Mon Jul 07 13:25:50 +0000 2025"	0	"NFT di sektor musik mulai dilirik musisi indie. Royalti otomatis t"	1942213471350976934			"id"				
4	1942195282894078159	"Mon Jul 07 12:13:34 +0000 2025"	0	"Bagaimana Nasib Royalti Musik Hamdan ATT? - https://t.co/VW"	1942195282894078159			"id"				
5	1942127696856903858	"Mon Jul 07 07:45:00 +0000 2025"	0	"NFT musik = royalti otomatis No lagi dibohongi label #Bybitindo"	1942127696856903858			"id"				
6	1942119638995001620	"Mon Jul 07 07:12:59 +0000 2025"	0	"NFT musik bakal meledak Bayangin punya royalti dari lagu viral"	1942119638995001620			"id"				
7	1942108262335389958	"Mon Jul 07 06:27:47 +0000 2025"	0	"NFT musik = cara baru musisi dapat royalti langsung dari fans. #"	1942108262335389958			"id"				
8	1942056302483849655	"Mon Jul 07 00:48:12 +0000 2025"	0	"10 Diskusi membahas sejumlah isu strategis antara lain posisi U"	1942056302483849655			"id"				
9	1941825325274726599	"Sun Jul 06 12:09:40 +0000 2025"	0	"Gue bersedia ngasih konsultasi/diskusi tentan"	1941825325274726599			"id"				
10	1941747303473348704	"Sun Jul 06 06:33:27 +0000 2025"	0	"NFT bisa jadi cara distribusi royalti musik yang lebih adil. #Bybit"	1941747303473348704			"id"				
11	1941404209892774123	"Sat Jul 05 07:50:07 +0000 2025"	0	"Bagaimana Nasib Royalti Musik Hamdan ATT? https://t.co/Zzcfah"	1941404209892774123			"id"				
12	1941151935308222847	"Fri Jul 04 15:07:40 +0000 2025"	0	"NFT musik terbukti jadi solusi untuk royalti transparan! #Bybitindo"	1941151935308222847			"id"				
13	1941116313122074710	"Fri Jul 04 12:46:07 +0000 2025"	0	"Serius deh ini store tempatku kerja udah ga bolehin play musik dar"	1941116313122074710			"id"				
14	1941019741164011752	"Fri Jul 04 06:22:23 +0000 2025"	1	"Polarisasi antara pencipta lagu dan penyanyi masih menjadi subje"	1941019741164011752			"id"				
15	1940938139130122627	"Fri Jul 04 01:37:57 +0000 2025"	0	"@liger_sjha @dearwono0 @liger_sjha Setelah investigasi perkin"	1940938139130122627			"id"				
16	1940750370638385456	"Thu Jul 03 12:32:00 +0000 2025"	0	"Kata Mutia Ayu mendiang Glenn Fredly menginginkan regenerasi"	1940750370638385456			"id"				

Gambar 1. Sample dataset

Berdasarkan Gambar 1, hasil crawling data dari Twitter (X) ditampilkan dalam format CSV yang memuat teks cuitan dan metadata pendukung. Sebagai contoh, salah satu entri berisi cuitan "NFT musik mulai ada aset royalti digital..." yang mencerminkan opini publik terkait isu royalti musik. Setiap data dilengkapi informasi waktu unggahan dan identitas cuitan, sehingga dataset siap digunakan pada tahap prapemrosesan dan analisis sentimen.

4.2. Pembersihan Data (Cleaning)

Tabel 1. Cleaning

Before	After
@detikcom Jadikan Indonesia negara tanpa musik...!! Awass...@ Nyanyi lagu ² kebangsaan juga kena royalti..!	Jadikan Indonesia negara tanpa musik Awass Nyanyi lagu kebangsaan juga kena royalti
@GoodRecom Lah kalo yg dari luar negeri apa lmk bayar royalti ke pemilik musik. Penipu	Lah kalo yg dari luar negeri apa lmk bayar royalti ke pemilik musik Penipu

Tabel 1 menampilkan contoh hasil tahap *cleaning* data pada tweet, yaitu penghapusan elemen non-tekstual seperti mention, simbol, dan tanda baca berlebih. Proses ini membuat teks menjadi lebih bersih, konsisten, dan siap digunakan pada tahap prapemrosesan selanjutnya.

4.3. Casefolding

Tabel 2. Case Folding

Before	After
Jadikan Indonesia negara tanpa musik Awass Nyanyi lagu kebangsaan juga kena royalti	jadikan indonesia negara tanpa musik awass nyanyi lagu kebangsaan juga kena royalti
Halo Shalihin Nad jangan khawatir Royalti musik hanya dikenakan untuk penggunaan komersial di ruang publik seperti kafe	halo shalihin nad jangan khawatir royalti musik hanya dikenakan untuk penggunaan komersial di ruang publik seperti kafe

Tabel 2 menampilkan contoh hasil penerapan tahap *case folding* pada data tweet. Pada proses ini, seluruh karakter alfabet pada teks diubah menjadi huruf kecil, sehingga perbedaan penulisan akibat penggunaan huruf kapital dapat dihilangkan. Perubahan ini bertujuan menjaga konsistensi representasi kata dan mendukung keakuratan proses pembentukan fitur pada tahap analisis selanjutnya.

4.4. Normalization

Tabel 3. Normalization

Before	After
murottal bacaan ngaji kan kok jd royalti musik ga sekalian adzan jg	murottal bacaan mengaji kan kok jadi royalti musik tidak sekalian adzan juga
tu bakal kena royalti ngga karena ada musik yang di setel	itu bakal kena royalti tidak karena ada musik yang di setel

Tabel 3 memperlihatkan hasil penerapan tahap *normalization* pada data tweet. Pada tahap ini, kata-kata dengan ejaan tidak baku, singkatan, serta bentuk bahasa informal diubah ke dalam bentuk baku. Perubahan tersebut bertujuan meningkatkan konsistensi bahasa pada data teks, sehingga mempermudah proses pengolahan dan pembentukan fitur pada tahap analisis sentimen selanjutnya.

4.5. Tokenizing

Tabel 4. Tokenization

Before	After
murottal bacaan mengaji kan kok jadi royalti musik tidak sekalian adzan juga	['murottal', 'bacaan', 'mengaji', 'kan', 'kok', 'jadi', 'royalti', 'musik', 'tidak', 'sekalian', 'adzan', 'juga']
itu bakal kena royalti tidak karena ada musik yang di setel	['itu', 'bakal', 'kena', 'royalti', 'tidak', 'karena', 'ada', 'musik', 'yang', 'di', 'setel']

Tabel 4 menampilkan contoh hasil penerapan tahap *tokenization* pada data tweet. Pada proses ini, teks yang semula berbentuk kalimat diuraikan menjadi kumpulan token berupa kata-kata tunggal. Pemecahan teks ini memungkinkan setiap kata diperlakukan sebagai unit analisis, sehingga mendukung tahapan prapemrosesan lanjutan dan pembentukan fitur dalam analisis sentimen.

4.6. Stopword Removal

Tabel 5. Stopword Removal

Before	After
['murottal', 'bacaan', 'mengaji', 'kan', 'kok', 'jadi', 'royalti', 'musik', 'tidak', 'sekalian', 'adzan', 'juga']	["murottal", "bacaan", "mengaji", "royalti", "musik", "sekalian", "adzan"]
['itu', 'bakal', 'kena', 'royalti', 'tidak', 'karena', 'ada', 'musik', 'yang', 'di', 'setel']	["bakal", "kena", "royalti", "musik", "setel"]

Tabel 5 menunjukkan hasil penerapan tahap *stopword removal* pada data tweet. Pada tahap ini, kata-kata umum yang tidak memiliki nilai informatif tinggi, seperti kata hubung dan kata keterangan, dihilangkan dari hasil tokenisasi. Proses ini bertujuan memperkecil dimensi data serta meningkatkan fokus analisis pada kata-kata yang lebih relevan dalam penentuan sentimen.

4.7. Stemming

Tabel 6. Stemming

Before	After
['murottal', 'bacaan', 'mengaji', 'kan', 'kok', 'jadi', 'royalti', 'musik', 'tidak', 'sekalian', 'adzan', 'juga']	["murottal", "bacaan", "mengaji", "royalti", "musik", "sekalian", "adzan"]
['itu', 'bakal', 'kena', 'royalti', 'tidak', 'karena', 'ada', 'musik', 'yang', 'di', 'setel']	["bakal", "kena", "royalti", "musik", "setel"]

Tabel 6 menampilkan contoh hasil penerapan tahap *stemming* pada data tweet. Pada proses ini, kata-kata yang memiliki imbuhan diubah ke dalam bentuk kata dasar sehingga variasi kata dengan makna yang sama dapat diseragamkan. Penerapan stemming bertujuan meningkatkan konsistensi representasi teks dan mendukung efektivitas pembobotan fitur serta proses klasifikasi sentimen pada tahap selanjutnya.

4.8. Labeling

Tabel 7. *Labeling*

Text	Labeling
bangga rakyat indonesia koruptor banyak koruptor diksh abolisi amnesti beras surplus harga beras mahal pasar tambah rakyat negara nikmat surplus beras indonesia rakyat serba dipajakin bahkan putar musik kena royalti	negatif
kadang polemik dunia musik soal karya soal atur beres sekarang benah musik indonesia biar sistem royalti lebih adil transparan musisi cipta laku usaha tenang benah musik indonesia	positif
kanwil kemenkum jabar diskusi atur kena kelola royalti musik kemenkum jabar asep sutandar	netral

Tabel 7 menunjukkan hasil pelabelan sentimen pada data tweet setelah melalui tahap prapemrosesan. Setiap teks diklasifikasikan ke dalam tiga kelas sentimen, yaitu positif, negatif, dan netral, berdasarkan kecenderungan polaritas yang terkandung di dalamnya. Hasil pelabelan ini digunakan sebagai data latih dalam proses pembentukan dan evaluasi model klasifikasi sentimen.

4.9. *TF-IDF*

Tabel 8. *TF-IDF*

Word	In documents	In Class Positif	In Class Netral	In Class Negatif
royalti	3674	2358	721	1202
musik	3584	2499	664	1059
bayar	1064	628	225	395
cipta	682	622	93	146
indonesia	717	505	117	163

Tabel 8 menampilkan contoh hasil representasi fitur menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)* pada data tweet. Tabel tersebut menunjukkan distribusi kemunculan beberapa kata kunci utama, seperti royalti, musik, dan bayar, baik secara keseluruhan dokumen maupun pada masing-masing kelas sentimen (positif, netral, dan negatif). Sebagai contoh, kata royalti dan musik memiliki frekuensi kemunculan tertinggi pada kelas positif, yang mengindikasikan dominasi pembahasan terkait kebijakan royalti musik pada sentimen tersebut. Informasi ini menunjukkan bahwa *TF-IDF* mampu menangkap kata-kata yang memiliki peran penting dalam membedakan kecenderungan sentimen dan digunakan sebagai dasar pembentukan fitur pada proses klasifikasi selanjutnya.

4.10. Support Vector Machine

Gambar 2 mengilustrasikan hasil implementasi algoritma *Support Vector Machine* (SVM) dalam proses klasifikasi sentimen pada data tweet. Pada tahap ini, data teks yang telah melalui prapemrosesan dan representasi fitur menggunakan *TF-IDF* diklasifikasikan ke dalam tiga kelas sentimen, yaitu

positif, negatif, dan netral. Setiap baris teks pada gambar menunjukkan hasil prediksi sentimen yang dihasilkan oleh model SVM, yang merefleksikan kemampuan model dalam memisahkan kelas sentimen berdasarkan pola fitur yang terbentuk pada ruang vektor berdimensi tinggi

teks	label_asli	
bayar royalti musik	positif	positif
sistem royalti kelas kuat semangat benah musik indonesia	netral	positif
makin masuk negeri rakyat kecil cari nafkah bayar royalti bidang musik bidang tani rakyat		
sulit pukup harga jual panen anjlok bidang dapur langka main elpoji begini dinergi kara raya		
rakyat kecil	negatif	negatif
Imak bego tuju bagus segala punya dimainin ipnya kudu bayar royalti punya masalah		
Imknnya goblok otak pakai bikin atur regulasi dulu biar semua enak bego cuma pungli		
misal pakai musik prem	negatif	negatif
nyala masalah royalti musik memang negeri doang cari masalah jela unfaedah stupid	negatif	negatif
bagaimana tukang cukur kalau potong rambut selalu full musik masa kena royalti	negatif	netral
pajak naik kur bpik tambah royalti musik tiba tiba dengan tunjangan anggota	negatif	
memang perintah indark pakai otak buat bijak	negatif	negatif
awal mula begini perkara royalti musik	negatif	negatif
benar rewel distribusi royalti putar tempat umum karya gara gara kalimat musik haram sering		
kali ucap anto lontar baji orang alas apa padahal anjur	negatif	negatif

Gambar 2. Hasil Prediksi SVM

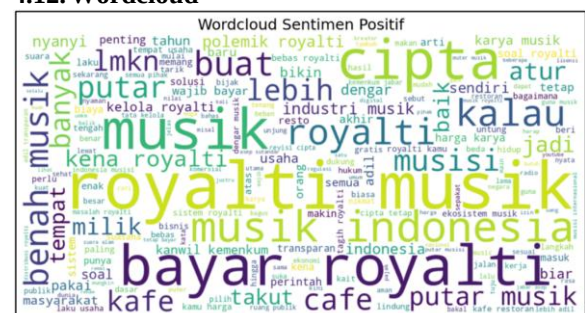
4.11. Confusion Matrix

	precision	recall	f1-score	support
negatif	0.77	0.80	0.79	311
netral	0.79	0.77	0.78	342
positif	0.52	0.49	0.50	110
accuracy			0.74	763
macro avg	0.69	0.69	0.69	763
weighted avg	0.74	0.74	0.74	763

Gambar 3. *Confusion Matrix*

Gambar 3 menunjukkan hasil evaluasi kinerja model klasifikasi sentimen menggunakan algoritma *Support Vector Machine (SVM)* yang disajikan dalam bentuk metrik *precision*, *recall*, dan *F1-score* untuk setiap kelas sentimen. Hasil evaluasi menunjukkan bahwa model memiliki kinerja yang relatif baik pada kelas negatif dan netral, dengan nilai *F1-score* masing-masing sebesar 0,79 dan 0,78, sedangkan kelas positif menunjukkan kinerja yang lebih rendah dengan *F1-score* sebesar 0,50. Secara keseluruhan, model mencapai tingkat akurasi sebesar 0,74, yang mengindikasikan bahwa *SVM* mampu mengklasifikasikan sentimen publik terhadap kebijakan royalti musik secara cukup andal, meskipun masih terdapat tantangan pada klasifikasi kelas positif akibat ketidakseimbangan distribusi data.

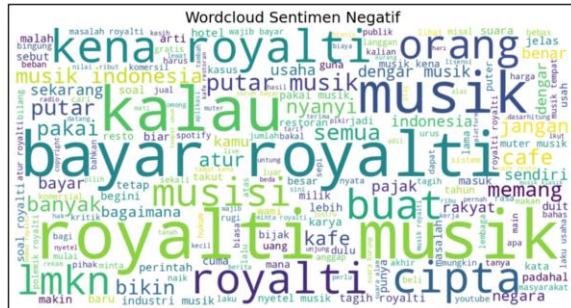
4.12. Wordcloud



Gambar 4. *Wordcloud* Sentimen Positif

Gambar 4 menunjukkan bahwa sentimen positif didominasi oleh kata-kata seperti royalti, musik, cipta, dan bayar, yang merefleksikan dukungan publik

terhadap perlindungan hak cipta dan kewajiban pembayaran royalti. Kemunculan istilah Indonesia, industri, musisi, dan cafe mengindikasikan bahwa sentimen positif juga berkaitan dengan harapan terhadap tata kelola industri musik yang lebih adil dan berkelanjutan dalam konteks pemanfaatan publik.



Gambar 5. *Wordcloud* Sentimen Negatif

Gambar 5 memperlihatkan bahwa sentimen negatif didominasi oleh kata-kata seperti royalti, bayar, kena, kafe, dan musisi, yang mencerminkan persepsi publik terhadap beban biaya dan kewajiban pembayaran royalti dalam pemanfaatan musik di ruang publik. Munculnya istilah aturan, LMKN, dan usaha menunjukkan bahwa sentimen negatif banyak diarahkan pada aspek regulasi dan implementasi kebijakan yang dinilai memberatkan pelaku usaha serta menimbulkan ketidakpastian dalam praktiknya.



Gambar 6. *Wordcloud* Sentimen Netral

Gambar 6 menunjukkan bahwa sentimen netral didominasi oleh kata-kata seperti royalti, musik, bayar, kelola, dan kanwil, yang mencerminkan pembahasan publik yang bersifat informatif dan deskriptif terkait pengelolaan serta kewajiban pembayaran royalti musik. Munculnya istilah LMKN, Kemenkum, hak cipta, dan lembaga menunjukkan bahwa sentimen netral banyak berfokus pada aspek regulasi, kelembagaan, dan mekanisme pelaksanaan kebijakan tanpa disertai penilaian emosional yang kuat.

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa penerapan teknik text mining dengan algoritma Support Vector Machine (SVM) efektif dalam mengklasifikasikan dan memetakan sentimen publik terhadap pelaksanaan Peraturan Pemerintah Nomor 56 Tahun 2021 tentang Pengelolaan Royalti Lagu dan/atau Musik di platform

Twitter. Hasil evaluasi model menggunakan confusion matrix menghasilkan akurasi sebesar 74,44% dengan kinerja yang relatif baik pada kelas negatif dan netral, meskipun performa pada kelas positif masih lebih rendah akibat ketidakseimbangan distribusi data. Analisis sentimen mengungkap bahwa opini publik didominasi oleh sentimen netral dan negatif, yang mencerminkan sikap informatif sekaligus kritis terhadap aspek implementasi kebijakan, sementara sentimen positif muncul dalam jumlah lebih terbatas dan berkaitan dengan dukungan terhadap perlindungan hak cipta dan tata kelola industri musik. Temuan ini menegaskan bahwa pendekatan SVM mampu memberikan gambaran yang komprehensif mengenai persepsi masyarakat terhadap kebijakan royalti musik. Untuk pengembangan penelitian selanjutnya, disarankan perluasan sumber data ke berbagai platform media sosial lain, penerapan algoritma alternatif berbasis deep learning, penambahan kategori sentimen yang lebih rinci, serta peningkatan kualitas dan kuantitas dataset agar analisis perubahan sentimen publik dapat dilakukan secara lebih mendalam dan akurat.

DAFTAR PUSTAKA

- [1] A. J. Walukow, "Pengelolaan royalti hak cipta lagu dan/atau musik menurut peraturan Pemerintah Nomor 56 Tahun 2021 tentang pengelolaan royalti hak cipta lagu dan/atau musik," *Lex Administratum*, vol. 10, no. 5, 2022.
- [2] A. L. Gunawan, E. Kuspraningrum, and F. N. Hediati, "Implementasi Penarikan Royalti Pengguna Lagu/Musik Pada Usaha Mikro, Kecil, Kafe di Kota Samarinda," *Jurnal Suara Hukum*, vol. 5, no. 1, pp. 190-206, 2023.
- [3] M. F. Rozi, N. K. Sari, and G. Puspaningrum, "Kepastian Hukum Penarikan Royalti Lagu dan/atau Musik Secara Komersial Terhadap Sektor Usaha Mikro, Kecil dan Menengah," *FAIRNESS AND JUSTICE: JURNAL ILMIAH ILMU HUKUM Yupedumenu: Universitas Muhammadiyah Jember*, vol. 21, no. 2, pp. 59-68, 2023.
- [4] F. S. Azzahra, N. S. Imaniyati, and R. Supriatna, "Aspek Hukum Pembayaran Royalti oleh Pemilik Kafe kepada Pencipta Musik Berdasarkan Peraturan Pemerintah Nomor 56 Tahun 2021 tentang Pengelolaan Royalti Lagu dan/atau Musik," in *Bandung Conference Series: Law Studies*, 2024, vol. 4, no. 2, pp. 1165-1170.
- [5] G. Indarsen, "Konsekuensi Hadirnya Peraturan Pemerintah Nomor 56 Tahun 2021 Tentang Pengelolaan Royalti Hak Cipta Lagu Dan/Atau Musik Terhadap Pemungutan Royalti Lagu Dan/Atau Musik," *Locus: Jurnal Konsep Ilmu Hukum*, vol. 3, no. 2, pp. 99-112, 2023.
- [6] D. A. Wiranti and P. A. F. Ariawantara, "Assessing the Role of Surabaya City Government's Responsiveness Toward Public Acceptance of COVID-19 Mitigation Policies,"

- Jurnal Borneo Administrator*, vol. 20, no. 1, pp. 85-100, 2024.
- [7] N. Syamsi, "Bahasa Tidak Resmi di Media Sosial: Analisis Komentar Netizen di Instagram Pusdatin Kemendikbudristek," *GERAM (Gerakan Aktif Menulis)*, vol. 12, no. 2, pp. 56-64, 2024.
- [8] A. Hermawan, I. Jowensen, and J. Junaedi, "Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine," *JST (Jurnal Sains Dan Teknologi)*, vol. 12, no. 1, pp. 129-137, 2023.
- [9] J. T. Kumalasari and I. Puspitorini, "Perbandingan Metode Klasifikasi dan SMOTE Terhadap Analisa Sentimen Mobil Listrik Indonesia," *Jurnal Minfo Polgan*, vol. 13, no. 2, pp. 2257-2268, 2024.
- [10] D. J. Haryanto, L. Muflikhah, and M. A. Fauzi, "Analisis sentimen review barang berbahasa Indonesia dengan metode support vector machine dan query expansion," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 9, pp. 2909-2916, 2018.
- [11] H. Y. Pradana, I. Slamet, and E. Zukhronah, "Analisis sentimen kinerja pemerintahan menggunakan algoritma nbc, knn, dan svm," *Prosiding Simposium Nasional Multidisiplin (SinaMu)*, vol. 4, pp. 114-121, 2023.
- [12] Y. A. Khawarga and Y. F. Riti, "Sentiment Analysis of Society Towards the Royalty Cases of Ahmad Dhani and Once Mekel Using Naïve Bayes," in *SINTEK-DC Seminar Nasional Teknologi Darma Cendika*, 2024, vol. 1, no. 1.
- [13] T. Safitri, Y. Umaidah, and I. Maulana, "Analisis Sentimen Pengguna Twitter Terhadap Grup Musik BTS Menggunakan Algoritma Support Vector Machine," *Journal of Applied Informatics and Computing*, vol. 7, no. 1, pp. 34-41, 2023.
- [14] S. Styawati, N. Hendrastuty, and A. R. Isnain, "Analisis sentimen masyarakat terhadap program kartu prakerja pada twitter dengan metode support vector machine," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 6, no. 3, pp. 150-155, 2021.
- [15] R. Ramlan, N. Satyahadewi, and W. Andani, "Analisis Sentimen Pengguna Twitter Menggunakan Support Vector Machine Pada Kasus Kenaikan Harga BBM," *Jambura Journal of Mathematics*, vol. 5, no. 2, pp. 431-445, 2023.
- [16] M. K. R. Rifa, M. H. Totohendarto, and M. R. Muttaqin, "Analisis Sentimen Pengguna E-Wallet Dana dan Gopay Pada Twitter Menggunakan Metode Support Vector Machine (SVM)," *TEKNIKA*, vol. 17, no. 2, pp. 323-332, 2023.
- [17] S. Agustina, E. Bonde, D. L. Salsabila, S. M. D. Hutabarat, and R. Wahyuni, "Perlindungan Hak Cipta Karya Musik di Dalam Digital Service Platform Berbasis User Generated Content Berdasarkan Undang-undang Hak Cipta," *Unes Journal of Swara Justisia*, vol. 8, no. 2, pp. 427-440, 2024.
- [18] M. Marali, P. N. Putri, and M. R. Alghifari, "Analisis Pelindungan Hukum Bagi Kreditor Pemegang Jaminan Hak Tanggungan Terhadap Pembatalan Sertifikat Hak Milik yang Sedang Dibebani Hak Tanggungan Oleh Pengadilan," *Padjadjaran Law Review*, vol. 10, no. 1, pp. 1-10, 2022.
- [19] Y. Arifardhani, "Problematisasi Lembaga Manajemen Kolektif Nasional (LMKN) Dalam Menghimpun Royalti Hak Cipta Di Indonesia," *SALAM: Jurnal Sosial dan Budaya Syar-i*, vol. 9, no. 3, pp. 865-872, 2022.
- [20] N. Muthmainnah, P. A. Pradita, and C. A. P. A. Bakar, "Perlindungan hukum terhadap hak cipta bidang lagu dan/atau musik berdasarkan PP Nomor 56 Tahun 2021 tentang pengelolaan royalti hak cipta lagu dan/atau musik," *Padjadjaran Law Review*, vol. 10, no. 1, pp. 110-123, 2022.
- [21] R. F. Kusumaningtyas, K. Roisah, D. Sulistyaningsih, and D. Aritenesa, "Pemenuhan Hak Pencipta Lagu oleh Pelaku Pertunjukan Tanpa Izin," *Jurnal Ilmiah Kebijakan Hukum*, vol. 19, no. 2, pp. 87-110, 2025.
- [22] T. Iswantono and A. F. Rosando, "Tinjauan yuridis pengelolaan royalti hak cipta lagu dan/atau musik menurut peraturan perundang-undangan yang berlaku di Indonesia," *Bureaucracy Journal: Indonesia Journal of Law and Social-Political Governance*, vol. 2, no. 1, pp. 108-118, 2022.
- [23] A. Juardi and M. Roestamy, "Analisis Hukum Terhadap Hak Ekonomi Pencipta Karya Musik Dan Lagu Yang Di Cover Version Pada Platform Digital," *Jurnal Ilmiah Living Law*, vol. 15, no. 2, pp. 129-140, 2023.
- [24] A. F. Rahman, "Klasifikasi Tweet di Twitter dengan Menggunakan Metode K-Nearest Neighbor," *Jurnal Sistim Informasi dan Teknologi*, pp. 64-69, 2022.
- [25] S. Mujahidin, M. N. Hasyim, and B. M. Pratama, "Implementasi Analisis Sentimen Opini Publik Mengenai Sirkuit Internasional Mandalika Pada Twitter Menggunakan Metode Multinomial Naïve Bayes Classifier," *Biaglala Informatika*, vol. 10, no. 2, pp. 129-136, 2022.
- [26] D. Rolliawati, K. Khalid, and I. S. Rozas, "Teknologi Opinion Mining untuk mendukung strategic planning," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 7, no. 2, pp. 293-302, 2020.
- [27] D. Y. Sihombing and Y. Nataliani, "Analisis Interaksi Pengguna Twitter pada Strategi Pengadaan Barang Menggunakan Social Network Analysis," *Sistemasi*, vol. 10, no. 2, pp. 434-444, 2021.