Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, Juni 2024 Page: 64−73
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5179

Analisis Sentimen Terhadap Publisher Rights Dalam Mengunggah Konten Digital Menggunakan Ensemble Learning

Anisa Putri*, Mustakim, Rice Novita, M Afdal

Sains dan Teknologi, Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
Email: [1,*]12050326104@students.uin-suska.ac.id, [2]mustakim@uin-suska.ac.id, [3]rice.novita@uin-suska.ac.id, [4]m.afdal@uin-suska.ac.id
Email Penulis Korespondensi: 12050326104@students.uin-suska.ac.id
Submitted: **15/05/2024**; Accepted: **23/06/2024**; Published**: 23/06/2024**

Abstrak−Konten digital mencakup berbagai bentuk informasi, mulai dari teks informatif hingga video interaktif. Dan YouTube, menjadi salah satu media sosial yang populer, menjadi platform paling banyak digunakan di Indonesia. Namun, kehadiran Rancangan Publisher Rights atau Rancangan Peraturan Presiden (Perpres) tentang Tanggung Jawab Platform Digital untuk Jurnalisme Berkualitas memunculkan perdebatan. Dalam konteks YouTube, peraturan ini berpotensi mengancam para konten kreator. Reaksi negatif dari berbagai pihak menunjukkan kekhawatiran terhadap dampak regulasi ini. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen terhadap Publisher Rights dalam pengunggahan konten digital menggunakan pendekatan ensemble learning. Ditemukan bahwa 60% sentimen adalah negatif, mencerminkan kekhawatiran terhadap hak cipta, royalti, atau isu etika. Sebanyak 32% sentimen adalah netral, menggambarkan ketidakpastian atau kurangnya informasi. Dan hanya 8% sentimen yang positif, mendukung kebijakan perlindungan hak penerbit dan mengakui nilai serta kontribusi mereka. Metode ini melibatkan teknik ensemble berbasis Bagging (Random Forest) dan Boosting (Adaboost), Dimana hasil akurasi dari Random Forest lebih tinggi dibandingkan dengan Adaboost, Random Forest dengan akurasi sebesar 83% sedangkan Adaboost dengan akurasi sebesar 68%.

Kata Kunci: Publisher Rights; Konten; Dewan Pers; Youtube; Ensemble Learning

Abstract−Digital content encompasses various forms of information, ranging from informative text to interactive videos. YouTube, as one of the most popular social media platforms, is widely used in Indonesia. However, the proposed Publisher Rights Bill or the Draft Presidential Regulation on the Responsibility of Digital Platforms for Quality Journalism has sparked debate. In the context of YouTube, this regulation has the potential to threaten content creators. Negative reactions from various parties highlight concerns about the impact of this regulation. Therefore, this study aims to analyze sentiment towards Publisher Rights in the uploading of digital content using an ensemble learning approach. The analysis found that 60% of the sentiment was negative, reflecting concerns about copyright, royalties, or ethical issues. A total of 32% of the sentiment was neutral, indicating uncertainty or a lack of information, and only 8% of the sentiment was positive, supporting the policy of protecting publisher rights and recognizing their value and contributions. This study employed ensemble techniques based on Bagging (Random Forest) and Boosting (Adaboost), where the accuracy of Random Forest was higher at 83% compared to Adaboost's accuracy of 68%.

Keywords: Publisher Rights; Content; Press Council; YouTube; Ensemble Learning

1. PENDAHULUAN

Saat ini Indonesia merupakan sebagai negara kedua terbanyak pengguna media sosial bersamaan dengan Brasil setelah India dengan rata-rata 8,4 jenis media sosial per bulannya. Seiring berkembangnya media sosial, konten digital menjadi pusat perhatian. Konten digital mencakup berbagai bentuk informasi, mulai dari teks informatif, gambar menarik, hingga video yang mendalam dan interaktif[1]. Melalui konten digital, kita tidak hanya menjadi pengamat, tetapi juga peserta aktif dalam masyarakat informasi ini. salah satu media sosial yang popular adalah *Youtube*[2], [3]. *Youtube* menjadi media sosial yang paling banyak di gunakan di Indonesia dengan total pengguna yaitu 139 juta di awal tahun 2023 kemudian diikuti *Facebook* dengan 19,9 juta pengguna[4], [5]. Dampak peggunaan *Youtube* ini memunculkan pengguna untuk memanfaatkan kreatifitas maupun mengekspresikan diri serta dapat berinteraksi dengan penontonnya sebagai salah satu fasilitas yang dimanfaatkan para kreator untuk mendapatkan keuntungan baik secara moril dan popularitas[6], [7]. Penghasilan seorang konten kreator bahkan termasuk penghasilan yang tinggi dari monetisasi atas perolehan dari viewer dan subscriber membuat banyak orang tertarik menjadi seorang konten kreator di *Youtube*[8], [9].

Namun muncul isu mengenai Rancangan Pepres Jurnalisme Berkualitas yang bertujuan untuk melindungi sumber daya keuangan media massa nasional, alasan terbentuknya aturan ini dikarenakan sekitar 60% belanja iklan masuk ke platform asing seperti *Google* dan *Facebook*. Namun rancangan ini memunculkan polemik dan regulasi yang dinilai berpotensi membawa kemunduran dunia media digital. Perpres ini juga dipadang menjadi ancaman bagi para konten kreator[10]. Karena dikhawatirkan dapat mematikan para kreator konten dan menguntungkan perusahaan pers secara sepihak, para kreator konten memiliki tiga kekhawatiran utama terhadap pengesahan Perpres *Publisher Rights*. Pertama, penghapusan konten berita yang tidak sesuai etika jurnalistik berdasarkan rekomendasi Dewan Pers dianggap dapat merugikan kreator dengan menenggelamkan konten mereka di bawah konten dari perusahaan pers, yang dinilai dapat merusak ekosistem media digital dan mengurangi netralitas berita. Kedua, sebagian kreator konten berpendapat bahwa Perpres *Publisher Rights* dapat menghilangkan peluang monetisasi dari konten yang mereka produksi. Ketiga, banyak kreator konten merasa keberatan dengan aturan yang mewajibkan pemberitahuan perubahan algoritma kepada perusahaan pers 28 hari sebelum perubahan tersebut dilakukan[11].

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, Juni 2024 Page: 64−73
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5179

Cepatnya penyebaran informasi tentang *Publisher Rights* ramai diperbincangkan, dan menimbulkan reaksi berupa dukungan, kekecewaan, bahkan penolakan dari para konten kreator di media sosial[9], [12]. Reaksi ini umumnya mengandung teks yang sering disebut sebagai sentimen, yang jika dianalisis dapat menghasilkan informasi yang berguna[13], [14]. Data teks tersebut dapat dianalisis menggunakan *text mining*, sebuah metode untuk mengekstrak informasi berkualitas tinggi dari kumpulan data yang tidak terstruktur, sehingga dapat mengidentifikasi masalah-masalah dalam teks pada topik tertentu dan dapat menemukan informasi penting dari sumber data dengan mengidentifikasi dan memeriksa pola tertentu [15].

Penelitian ini menggunakan model klasifikasi *Ensemble Bagging (Random Forest)* dan *Ensemble Boosting (Adaboost)*. Mengkombinasikan model klasifikasi untuk membentuk model baru atau mengoptimalkan model yang tersedia yang lebih dikenal sebagai metode *ensemble* juga dapat di terapkan untuk mengatasi masalah yang ditemui pada analisis sentimen mengingat metode *ensemble* cendrung lebih akuran dibandingkan dengan model klasifikasi utama atau pembentuknya[2].

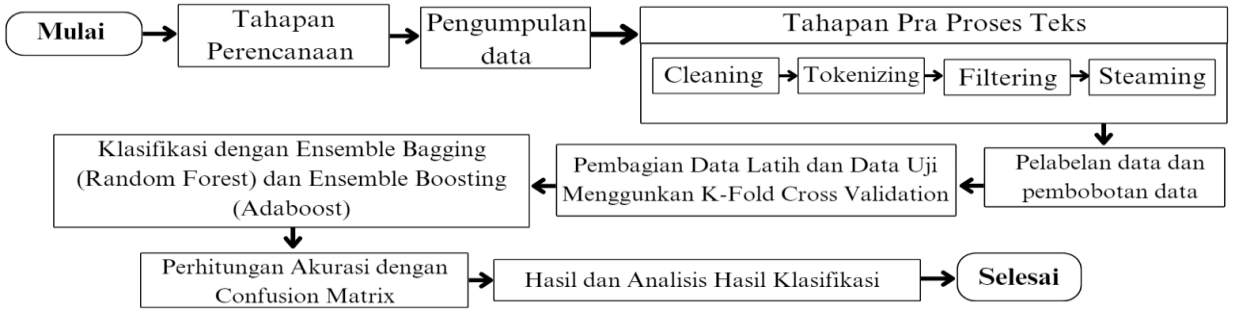
Berdasarkan penelitian yang dilakukan Chaudhary Jagrit Varshney dkk (2020), membandingkan teknik klasifikasi seperti *Random Forest*, *Support Vector Machine*, *Logistic Regression*, *Naïve Bayes*, dan SGD *Classifier* digunakan untuk mengklasifikasikan data, dan Pengklasifikasi ensemble (*Random Forest*) menemukan hasil optimal [16], adapun penelitian yang dilakukan Dimple Tiwari dkk, (2020) menilai kinerja teknik ansambel berbasis Bagging dan *Boosting* untuk SA jaringan sosial, menunjukkan bahwa teknik ansambel berbasis bagging mengungguli teknik berbasis *Boosting* untuk klasifikasi teks [2]. Lalu penelitian yang dilakukan oleh Sergio González, dkk (2020) Membandingkan algoritma *Bagging* dan *Boosting* yang paling terkenal dan perangkat lunaknya. Dan fokus pada 14 algoritma dan menyatakan *Ensemble Boosting* menunjukkan efisiensi yang lebih besar [17]. Ada juga penelitan dari I Bomoiye Domor Mienye and Yanxia Sun (2022) yang membedakan antara tiga metode ansambel utama yakni: bagging, boosting, dan stacking. Dan menyatakan bahwa Random Forest dan XGBoost paling banyak digunakan dalam literatur [18]. Dan Monika Kabir, dkk (2021) menyatakan metode pembelajaran mesin yang terkenal ditinjau dan dibandingkan satu sama lain. Metode terkenal seperti Support Vector Machine, Decision Tree, Bagging, Boosting, Random Forest dan Maximum Entropy diimplementasikan dalam eksperimen. Dapat disimpulkan Boosting berkinerja cukup baik [19].

Maka dalam penelitian ini akan menganalisis sentimen terhadap *publisher rights* dalam mengunggah konten digital menggunakan *Ensemble Learning*. dimana kategori *ensemble* yang di gunakan dalam penelitian ini adalah *Bagging (Random Forest)* dan *Boosting (Adaboost)* yang dimana untuk mengetahui kinerja dari Klasifikasi *Ensemble* yang digunakan untuk setiap fitur ekstraksi terhadap topik *Publisher Rights* Dalam Mengunggah Konten Digital yang mencakup respons positif, negatif, atau netral terhadap kebijakan, kedua dapat menjadi bahan evaluasi pemerintah dalam rancangan aturan *Publisher Rights* untuk memahami sikap dan pandangan publik sehingga mampu mengambil Tindakan yang sesuai untuk menyempurnakan kebijakan. Dan yag ketiga di harapkan dapat menjadi referensi untuk penelitian berikutnya mengenai analisis sentiment dari *social media youtube* untuk topik yang berbeda.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Metodologi penelitian dalam studi ini secara umum terdiri dari delapan tahapan. Berikut adalah tahapan-tahapan penelitian tersebut yang dapat dilihat pada Gambar 1.



Gambar 1. Metodologi Penelitian

Berikut dalam menyelesaikan penelitian ini dilakukan secara sistematis dengan tahapan-tahapan metodologi penelitian pada gambar 1 seperti berikut:

a. Perencanaan

Dalam tahap ini meliputi Observasi dan wawancara di Dinas Komunikasi Informatika Statistik dan Persandian Pemerintah Kota Pekanbaru dengan melakukan wawancara kepada ibu Verdhira Dinanti, S.I.Kom yang sekarang menjabat sebagai Pranata Humas Ahli Muda

b. Pengumpulan Data

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, Juni 2024 Page: 64−73
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5179

Menggunakan metode *scraping* dengan bahasa pemrogramman *Python* yaitu *youtube-coment-downloade*r pada data komentar *Youtube* diakun Channel yang mewakili Dewan Pers, Kominfo, dan Media Massa dengan kata kunci "*Publisher Right*" pada *Youtube*, dengan total 1917 data.

c. Pre-Processing

Pada tahap ini dilakukan *pre-processing* data karena data komentar YouTube yang dihasilkan tidak sistematis. Tahap *pre-processing* yang dilakukan adalah *cleaning*, *tokenizing*, *filtering*, dan *stemming*.

d. Pelabelan

Pada tahap ini, pelabelan kata dilakukan untuk mengidentifikasi sentimen positif, netral, dan negatif. Proses ini dilakukan secara manual oleh pakar bahasa, yaitu Elvina S.Pd.

e. Pembobotan TF-IDF

Digunakan teknik TF-IDF untuk melakukan pemrosesan kata dan mendapatkan bobot nilai untuk setiap kata.

f. Klasifikasi *Ensemble Learning*

Pada tahap klasifikasi dibagi menjadi dua pembagian data yaitu data training dan data testing dengan menggunakan metode perhitungan 10 *K-fold Cross Validation* dan *Ensemble Learning*. Dimana kategori *ensemble* yang di gunakan dalam penelitian ini adalah *Bagging (Random Forest) dan Boosting (Adaboost)*.

g. Visualisasi Kata dan Analisis Konteks

Tahap akhir dari penelitian ini adalah memperoleh visualisasi kata dengan website *wordclouds* dan Analisis konteks

2.2 *Pre-Processing Data*

Bertujuan untuk mengubah data yang tidak terstruktur menjadi bentuk yang siap untuk dianalisis. Beberapa tahapan *Pre Processing* yang diterapkan adalah memastikan apakah ada baris kosong pada dataset[20].Teks terkadang tidak dalam format yang dapat digunakan. Untuk mendapatkan teks dalam format yang dapat digunakan dengan berbagai metode pemrosesan awal membersihkan teks diterapkan.

Pre-Processing Data berisi empat langkah utama berikut sebelum meneruskan teks ke pengklasifikasi, yang pertama *Cleaning* ialah menghilangkan/menghapuskan karakter lain selain huruf, seperti menghapus *hashtag* (#), *link*, dan mengubah teks menjadi huruf kecil[21]. *Tokenizing* ialah teks dibagi menjadi beberapa token berdasarkan karakter pemisah seperti spasi, koma, tab, dll. Atau fase pemisahan mendasar dalam sentimen analisis yang mempartisi kalimat, frasa, atau paragraf menjadi satu kata yang disebut token[2], [22]. *Stop-words* atau *Filtering* kata-kata yang umum digunakan dan memiliki sedikit kandungan informasi seperti konjungsi, preposisi. *Stemming* adalah mereduksi kata menjadi batangnya dan menghilangkan akhiran kata tersebut, menurut beberapa aturan tata bahasa[22], [23].

2.3 Pembobotan TF -IDF (*Term Frequency-Inverse Document Frequency*)

Dalam penelitian ini *Term Frecuency-Inverse Document Frecuency* (TF-IDF) untuk menghitung bobot setiap kata yang diekstraksi untuk menghitung kata-kata umum dalam pencarian informasi[2], [23]. *Term Frequency* (TF) adalah metode penghitungan frekuensi kemunculan istilah dalam suatu dokumen, dan *Reverse Document Frequency* (IDF) adalah metode penghitungan istilah yang muncul dalam berbagai dokumen (komentar) yang dianggap istilah umum dianggap tidak penting[14][24].

2.4 Klasifikasi *Ensemble Learning*

Ensemble learning mengkombinasi beberapa pengklasifikasi untuk meningkatkan akurasi prediksi secara keseluruhan[17], [18]. Banyak algoritma ML konvensional cenderung berkinerja buruk ketika dilatih dengan kumpulan data yang tidak seimbang, Oleh karena itu, para peneliti sering kali menggunakan pendekatan pembelajaran yang baru dan lebih baik, seperti *Ensemble Learning*[18]. untuk menghasilkan klasifikasi, regresi, prediksi dan pemeringkatan yang lebih akurat dan andal dibandingkan dengan pengklasifikasi tunggal atau konvesional untuk generalisasi tanpa mengabaikan lebih banyak pengetahuan lokal atau spesifik Performanya yang baik telah menjadikan dirinya sebagai salah satu metode pembelajaran mesin terbaik dan paling berpengaruh[17]. Metode *Ensembel* yang dibagi menjadi 3 kategori utama: *Bagging*, *Boosting* dan *Stacking*[2], [13], [17]. Bagging dan *Boosting* adalah dua pendekatan utama pembelajaran *ensembel* yang berisi berbagai algoritma *ensembel* untuk mengklasifikasikan polaritas sentimen[2].

2.4.1 *Ensemble Bagging (Bootstrap Aggregating)*

Ensemble Bagging dikembangkan oleh Breiman tahun 1994 untuk meningkatkan kinerja klasifikasi model machine learning dengan menggabungkan prediksi dari set pelatihan yang dihasilkan secara acak[17], [18]. dengan penggantian (bootstrap) dari dataset pelatihan asli[19], [25], *Bagging* lebih meningkatkan kinerja pembelajar dasar jika algoritme yang digunakan dalam mempelajari model tidak stabil dan bekerja secara optimal ketika anggota *Ensemble* memiliki varian tinggi dan bias rendah, Pendekatan ini dapat mengurangi overfitting pada model yang kompleks[13].contohnya *decision tree*. dan algoritma yang di gunakan adalah *Random Forest*, *Extra Tree* dan *Meta Estimator* (*Linear* SVC) yang menggunakan sampel bootstrap[17], [18]. *Random Forest* sering dipilih untuk

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, Juni 2024 Page: 64−73
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5179

penelitian di antara algoritma *ensemble bagging*, karena mudah untuk diimplementasikan, cepat dan memperoleh kinerja yang sangat baik. Berikut adalah rumus matematika dan parameternya dari algoritma *Random Forest*[19],

$$F(x) = \frac{1}{J} \sum_{j=1}^{J} f_j(x) \quad (1)$$

Dalam rumus (1), F(x) merupakan Prediksi akhir yang dihasilkan oleh *Random Forest* untuk instance . $\frac{1}{J}$ merupakan rata-rata dari prediksi seluruh pohon keputusan. Ini mencerminkan prinsip "majority voting" atau rata-rata dalam kasus regresi. $\sum_{j=1}^{J}$ merupakan Simbol sigma ini menunjukkan bahwa kita sedang menjumlahkan prediksi dari semua pohon keputusan dalam ensemble. Dan f j (x) merupakan prediksi yang dihasilkan oleh pohon keputusan ke-j untuk instance x.

2.4.2 *Ensemble Bagging (Bootstrap Aggregating)*

Boosting merupakan pembelajaran *ensemble* yang baik untuk klasifikasi teks ,meningkatkan kinerja pelajar yang lemah dengan menjalankan beberapa subset kumpulan data secara berurutan[18]. untuk berubah menjadi pembelajar yang kuat dengan memberikan bobot. Dan dilatih untuk mengurangi varians dan mencapai kinerja yang lebih baik[2], [19]. *Boosting* berfokus pada pelatihan beberapa model secara berurutan, di mana setiap model berusaha untuk memperbaiki kesalahan prediksi model sebelumnya. Namun metode *Boosting* sangat sensitif terhadap noise[19]. Contoh Algoritma: *AdaBoost*, *Gradient Boosting* (misalnya, *XGBoost*, *LightGBM*, dan *CatBoost*) [20]. *AdaBoost* adalah yang paling populer. Algoritme *AdaBoost* cepat, sederhana, dan mudah dimodelkan dengan lebih sedikit kebutuhan penyetelan *hyperparameter*. Berikut adalah rumus matematika dan parameternya dari algoritma *Adaboost*[18].

$$H(x) = sign(\sum_{t=1}^{T} \alpha_t h_t(x)) \quad (2)$$

Dalam rumus (2), $H(x)$ merupakan prediksi akhiri untuk inputan x. $h_t(x)$ merupakan keluaran dari pengklasifikasi lemah t untuk masukan x . α_t merupakan bobot yang ditetapkan ke pengklasifikasi. T merupakan jumlah total model lemah yang dibentuk.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan data

Pada penelitian ini dilakukan dengan cara *scraping* pada *Youtube* menggunakan bahasa pemrogramman *Python*. Dan didapatkan hasil sebanyak 1917 data. Data yang digunakan dalam penelitian ini adalah data komentar dalam bahasa Indonesia dari aplikasi *Youtube*, dengan rentang waktu 01 Januari 2023 – 31 Oktober 2023, video berasal dari Kanal terverifikasi yang di filter berdasarkan jumlah tayangan per tanggal 14 November 2023 dengan kata kunci "*Publisher Rights*". Berikut data awal yang dapat dilihat pada Tabel 1.

Tabel 1. Data Awal

No	Teks
1	da tua mending pensiun boomer lg disrupsi otak mana da deh boomer gnti aja heran dipake maaf ye idiot
2	Bu kok videonya ada iklan ya?
3	Kalo bisa sih google mending angkat kaki aja dari Indonesia, termasuk semua layanannya, Gmail, android, youtube dan seluruhnya. Biar adu kuat aja pada. Ini kalo gw jadi CEO Google main kuat kuatan aja lu GK pake android. Kita liat aja siapa yg akhir nya ngemis ngemis. Suru dewan pers aja dia bikin google tandingan, sama operating sistem pengganti android
4	DEWAN PERS HARUS RAJIN, MENGAPA KOALISI WARTAWAN RANGKING INDONESIA(KWRI), tidak dimasukan di google, Mengapa?
…	…
1917	Owi owi gblk

3.2 Tahap *Preprocessing Data*

Beberapa tahapan preprocessing yang dilakukan meliputi pembersihan data dengan menghapus angka, simbol, emoticon, dan URL. Kemudian dilakukan seleksi data yang relevan dengan topik penelitian. Selanjutnya, proses *tokenizing* dilakukan untuk memecah kalimat menjadi kata-kata, diikuti oleh penghapusan *stopword* dan *filtering* , yaitu kata-kata yang tidak memiliki nilai positif atau negatif. Terakhir, dilakukan *stemming* untuk mengubah kata berimbuhan menjadi bentuk dasarnya, berikut *Preprocessing Data* yang dapat dilihat pada Tabel 2.

Tabel 2. Data Hasil Preprocessing

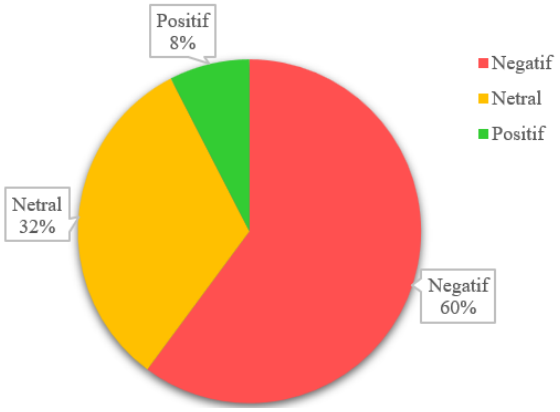
No	**Teks**
1	da tua mending pensiun boomer lg disrupsi otak mana da deh boomer gnti aja heran dipake maaf ye idiot

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, Juni 2024 Page: 64−73
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5179

2	bu video iklan
3	sih google mending angkat kaki aja indonesia layan gmail android youtube biar adu kuat aja gw ceo google main kuat kuat aja pake android liat aja ngemis ngemis suru
4	dewan pers rajin koalisi wartawan rangking indonesia kwri masuk google
…	
1917	owi owi gblk

3.3 Tahap Pelabelan Data

Data komentar yang diperoleh tidak memiliki kelas/label sehingga diperlukan sebuah metode untuk melakukan pelabelan. Pada penelitian ini dilakukan pelabelan sentiment secara manual menjadi 3 tiga kelas , yaitu positif, negatif dan netral, dengan jumlah data tweet sebanyak 1917 data oleh pakar yang memiliki latar belakang sebagai Dosen Bahasa Indonesia. Hasil pelabelan data dapat dilihat pada Gambar 2 dibawah ini.



Gambar 2. Hasil pelabelan

Berdasarkan hasil analisis sentimen terhadap topik "PUBLISHER RIGHTS DALAM MENGUNGGAH KONTEN DIGITAL", ditemukan bahwa 60% sentimen dikategorikan sebagai negatif yang mencerminkan kekhawatiran atau ketidakpuasan terhadap implementasi Publisher Rights, terkait dengan masalah hak cipta, pembagian royalti, atau isu-etika lainnya yakni sebanyak 1153 sentiment. Lalu 32% sentiment dikategorikan sebagai netral yang menggambarkan kelompok yang tidak memiliki sikap ekstrem, atau mungkin menunjukkan ketidakpastian atau kurangnya informasi yang memadai dalam pandangan mereka terkait topik ini yakni sebanyak 619 sentimen. Selain itu terdapat 8% sentiment dikategorikan positif yang mencerminkan dukungan terhadap kebijakan yang melindungi hak-hak penerbit atau pengakuan terhadap nilai dan kontribusi mereka yakni sebanyak 145 sentiment.

Dengan memperhatikan analisis sentimen ini, pemangku kepentingan dapat memahami lebih baik dinamika opini publik terkait hak-hak penerbit, memperbaiki atau menyesuaikan regulasi dan merumuskan langkah-langkah yang sesuai untuk memenuhi kebutuhan dan mengatasi isu isu yang di hadapi media dan masyarakat. hal ini juga mambantu media dan konten kreator untuk menyesuaikan strategi dalam menghadapi regulasi sehingga mampu menciptakan konten yang lebih sesuai dengan harapan masyarakat. Data pelabelan dapat dilihat pada Tabel 3 dibawah ini.

Tabel 3. Data Pelabelan

No	Teks	Pelabelan
1	da tua mending pensiun boomer lg disrupsi otak mana da deh boomer gnti aja heran dipake maaf ye idiot	negatif
2	bu video iklan	netral
3	sih google mending angkat kaki aja indonesia layan gmail android youtube biar adu kuat aja gw ceo google main kuat kuat aja pake android liat aja ngemis ngemis suru	netral
4	dewan pers rajin koalisi wartawan rangking indonesia kwri masuk google	negatif
...	...	...
1917	Owi owi	netral

3.4 Pembobotan Data TF-IDF

TF-IDF dilakukan untuk mendapatkan nilai atau bobot dari suatu kata yang telah diekstrak. Metode ini banyak diterapkan dalam text retrieval dan text preprocessing. Perhitungan dilakukan menggunakan python dengan memakai modul skicit-learn. TF-IDF dibagi sesuai dengan data training dan data set. Hasil Data TF-IDF dapat dilihat pada Tabel 4 dibawah ini.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, Juni 2024 Page: 64−73
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5179

Tabel 4. Hasil TF-IDF

No	Term								
	abcd	ada	adalah	adu	aja	akhir	...	youtube	zoomer
1	0.0	0.0	0.18	0.0	0.09	0.0	...	0.0	0.0
2	0.0	0.25	0.0	0.0	0.0	0.0	...	0.0	0.0
3	0.0	0.0	0.0	0.14	0.30	0.14	...	0.08	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0
...	...	...	...	...	...	...	...	...	...
1917	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0

3.5 Pembagian Data

Berdasarkan dengan metodolgi penelitian yang telah dijelaskan, Distribusi data pelatihan dan pengujian berdasarkan metode yang dijelaskan dilakukan menggunakan *K-Fold Cross Validation* pada data tweet *preprocessing* dan TF-IDF. Kemudian akan diproses dengan algoritma *Random Forest* dan *Boosting*.

3.5.1 Perhitungan *Ensemble Bagging*

Dalam penelitian ini algoritma *Random Forest* (RF) menunjukkan bahwa parameter terbaik untuk model *Random Forest* setelah melakukan *Grid Search* adalah {*'max_depth': 30, 'n_estimators': 10*}. Artinya, model yang memiliki kedalaman maksimum sebanyak 30 dan jumlah pohon sebanyak 10 memberikan performa yang optimal berdasarkan metrik evaluasi akurasi. Berikut gambaran rinci tentang kinerja model *Random Forest.* Hasil Klasifikasi *Ensemble Bagging* dapat dilihat pada Tabel 5 dibawah ini.

Tabel 5. Hasil Klasifikasi *Ensemble Bagging*

Class	**Precision %**	**Recall %**	**F1-Score %**
Negatif	81	91	86
Netral	75	84	79
Positif	98	66	79

3.5.2 Perhitungan *Ensemble Boosting*

Dalam penelitian ini Algoritma *Adaboost* menunjukkan bahwa parameter terbaik untuk model *AdaBoost* setelah melakukan *Grid Search* adalah {*'learning_rate': 1.0, 'n_estimators': 50*}. Artinya, model yang memiliki learning rate sebanyak 1.0 dan jumlah estimator (pohon kecil) sebanyak 50 memberikan performa yang optimal berdasarkan metrik evaluasi akurasi. Berikut gambaran rinci tentang kinerja model *Adaboost* pada data pelatihan. Hasil Klasifikasi *Ensemble Bosting* dapat dilihat pada Tabel 6 dibawah ini.

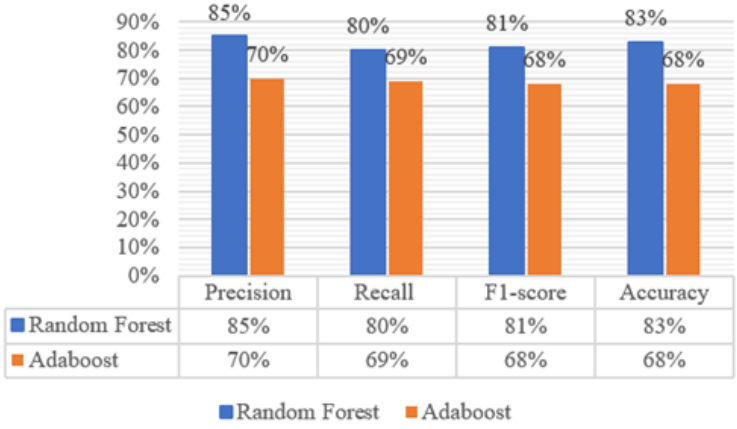
Tabel 6. Hasil Klasifikasi Ensemble Boosting

Class	**Precision %**	**Recall %**	**F1-Score %**
Negatif	76	66	70
Netral	55	85	67
Positif	79	57	66

3.6 Analisis Klasifikasi

Percobaan memakai metode pembagian data dengan *K-Fold Cross Validation* dimana didapat hasil perbandingan akurasi antara *Random Forest* dan *Adaboost* terhadap data Komentar di *youtube*. Dimana hasil akurasi dari *Random Forest* lebih tinggi dibandingkan dengan *Adaboost*, *Random Forest* dengan akurasi sebesar 83% sedangkan *Adaboost* dengan akurasi sebesar 68%. Dan menggunkan nilai *Macro avg*, yakni rata-rata dari setiap metrik evaluasi (*precision, recall, dan F1-score*) untuk setiap kelas, dihitung secara keseluruhan. Hasil Perbandingan Klasifikasi *Ensemble Bagging (Random Forest)* dan *Ensemble Boosting (Adaboost)* dapat dilihat pada gambar 3 dibawah ini.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, Juni 2024 Page: 64−73
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5179



Gambar 3. Perbandingan Performa *Random Forest* dan *Adaboost*

3.7 Visualisasi Data

Untuk memahami kata-kata yang sering muncul dalam komentar data berdasarkan kelasnya, visualisasi data dilakukan menggunakan *word cloud*. Ukuran font kata dalam *word cloud* menunjukkan frekuensi kemunculan kata tersebut. Semakin besar ukuran *font*, semakin tinggi frekuensi kemunculan kata tersebut. Dengan memanfaatkan situs web yakni *https://voyant-tools.org/* [26].

3.7.1 Visualisai Data Positif

Kemunculan dengan 3 frekuensi tertinggi pada opini positif pada data teks mengenai *Publisher Right* yaitu “Media”, “Konten” dan ”Kreator”. Berikut adalah visualisasi data yang positif menggunakan *wordcloud* dapat dilihat pada gambar 4.

Gambar 4. Visualisasi Data Positif

3.7.2 Visualisasi Data Negatif

Kemunculan dengan 3 frekuensi tertinggi pada opini negatif pada data teks mengenai *Publisher Right* yaitu “Dewan”, ”Pers” ,”Bubar” dan ”Demo”. Berikut adalah visualisasi data yang positif menggunakan *wordcloud* dapat dilihat pada gambar 5.

Gambar 5. Visualisasi Data Negatif

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, Juni 2024 Page: 64−73
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5179

3.7.3 Visualisasi Data Netral

Kemunculan dengan 3 frekuensi tertinggi pada opini netral pada data teks mengenai *Publiser Right* yaitu “Atur”, “Negara” dan ”Indonesia”. Berikut adalah visualisasi data yang netral menggunakan *wordcloud* dapat dilihat pada gambar 6.

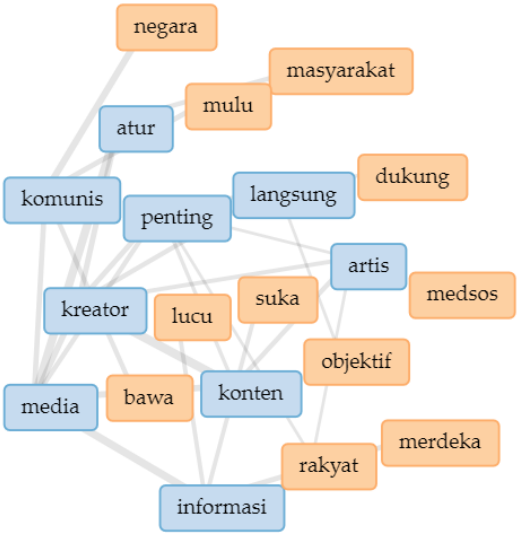
Gambar 6. Visualisasi Data Netral

3.8 Analisis Konteks

Selanjutnya peneliti ini akan mencoba mendapatkan konteks dari data yang telah diperoleh. Alat untuk memvisualisasikan konteks ini adalah *wordlink*. *WordBridge* atau *Wordlink* adalah metode visualisasi berbasis grafik inovatif yang bertujuan untuk mengilustrasikan hubungan antar entitas dalam korpora tekstual. Berfungsi untuk melihat kata terkait dari setiap teks yang muncul pada data. Pada *wordlink*, kata yang berlabel biru merupakan kata utama sedangkan kata yang berlabel oren merupakan kata sekunder atau kata yang sering dipasangkan dengan kata utama. Selain itu perbedaan ketebalan garis juga mempunyai arti tersendiri, jika garis penghubung antar kata tebal maka kata-kata tersebut sering muncul dalam kalimat yang sama, sebaliknya jika garis penghubung tipis maka kata-kata sambung tersebut muncul lebih sedikit dalam kalimat yang sama[27].

3.8.1 Konteks Data Positif

Berdasarkan visualisasi pada Gambar 7, terlihat adanya kata utama seperti ”media” dan kata skunder yang sering di pasangkan dengan kata utama seperti ”negara”, dilihat berdasarkan warna. Pada wordlink data positif frasa dalam teks terkait lebih bervariasi. Hal ini didukung dengan adanya garis tipis yang menghubungkan kata-kata, artinya kata-kata yang muncul tidak sering dalam satu kalimat atau kata sering muncul dalam kalimat yang berbeda. kata yang bermunculan memuliki jumlah yang tidak jauh berbeda di lihat dari ketebalan suatu kata yang memiliki ukuran yang sama .

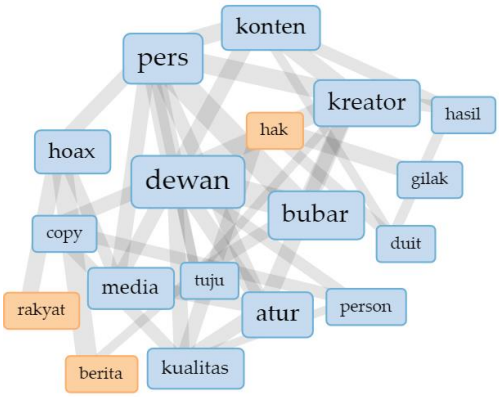


Gambar 7. Konteks Data Positif

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, Juni 2024 Page: 64−73
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5179

3.8.2 Konteks Data Negatif

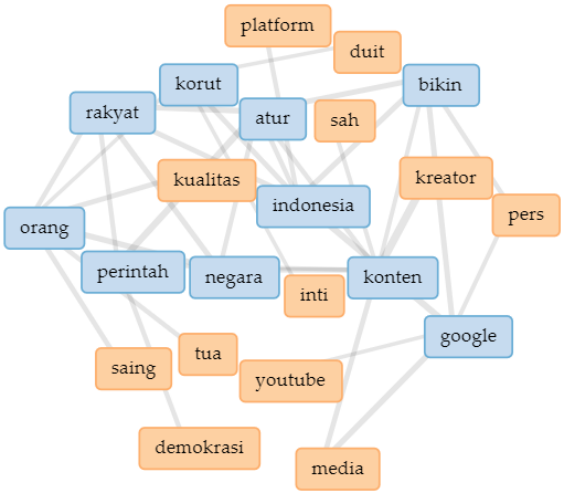
Berdasarkan visualisasi pada Gambar 8, terlihat adanya kata utama seperti ”dewan” dan kata skunder yang sering di pasangkan dengan kata utama seperti ”rakyat”, dilihat berdasarkan warna. Pada wordlink data negatif frasa dalam teks terkait tidak terlalu bervariasi. Hal ini didukung dengan adanya garis tebal yang menghubungkan kata-kata, artinya kata-kata yang muncul sering dalam satu kalimat. kata yang bermunculan memuliki jumlah yang berbeda di lihat dari ketebalan suatu kata yang memiliki ukuran yang yang berbeda. Dan terdapat kata yang sangat dominan yakni ”Dewan” terlihat bahwa isu ”Dewan” yang selalu ramai diperbincangkan.



Gambar 8. Konteks Data Negatif

3.8.3 Konteks Data Netral

Berdasarkan visualisasi pada Gambar 9, terlihat adanya kata utama seperti ”Negara” dan kata skunder yang sering di pasangkan dengan kata utama seperti ”Kreator”, dilihat berdasarkan warna. Pada wordlink data positif frasa dalam teks terkait lebih bervariasi. Hal ini didukung dengan adanya garis tipis yang menghubungkan kata-kata, artinya kata-kata yang muncul tidak sering dalam satu kalimat atau kata sering muncul dalam kalimat yang berbeda. kata yang bermunculan memuliki jumlah yang tidak jauh berbeda di lihat dari ketebalan suatu kata yang memiliki ukuran yang sama.



Gambar 9. Konteks Data Netral

4. KESIMPULAN

Berdasarkan dari rangkaian setiap proses analisi sentiment yang dilakukan menggunakan bahasa pyhton dengan 1917 data mengenai Publisher Right, ditemukan bahwa 60% sentimen (1153 sentimen) adalah negatif, mencerminkan kekhawatiran terhadap hak cipta, royalti, atau isu etika. Sebanyak 32% sentimen (619 sentimen) adalah netral, menggambarkan ketidakpastian atau kurangnya informasi. Dan hanya 8% sentimen (145 sentimen) yang positif, mendukung kebijakan perlindungan hak penerbit dan mengakui nilai serta kontribusi mereka. Lalu diperoleh hasil perbandingan yakni *Random Forest* memberikan performa yang lebih baik pada data pelatihan mencapai akurasi 83% dan memiliki metrik evaluasi kelas yang lebih baik secara keseluruhan dibandingkan dengan *AdaBoost* yakni 68%. Didapatkan juga kemunculan kata dengan 3 frekuensi tertinggi pada data positif yakni “media”,”konten”,”kreator”, pada data negatif yakni “dewan”,’pers”,”bubar”. Dan pada data netral yakni “atur”,”negara”,”Indonesia”.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, Juni 2024 Page: 64−73
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5179

REFERENCES

[1] O. M. Hylland, H. Stavrum, M. T. Heian, B. Kleppe, and K. P. Miland, "Creating careers in the kingdom of content. The platform-dependence and platform-ambivalence of digital cultural labour in Norway," *Poetics*, vol. 103, no. February, p. 101885, 2024, doi: 10.1016/j.poetic.2024.101885.

[2] D. Tiwari, B. Nagpal, B. S. Bhati, A. Mishra, and M. Kumar, *A systematic review of social network sentiment analysis with comparative study of ensemble-based techniques*, vol. 56, no. 11. Springer Netherlands, 2023. doi: 10.1007/s10462-023-10472-w.

[3] A. Z. Yonatan, "10 Negara dengan Pengguna Jenis Media Sosial Terbanyak 2023 - GoodStats Data," *GoodStats*, 2023.

[4] A. Z. Yonatan, "Menilik Pengguna Media Sosial Indonesia 2017-2026 - GoodStats Data," *GoodStats*, 2023.

[5] E. F. Setiadi, A. Azmi, and J. Indrawadi, "Youtube Sebagai Sumber Belajar Generasi Milenial," vol. 2, no. 4, pp. 313–323, 2019.

[6] B. S. Chai, T. Chae, and A. L. Huang, "Evaluation of Educational YouTube Videos for Distal Radius Fracture Treatment," *J. Hand Surg. Glob. Online*, pp. 1–6, 2024, doi: 10.1016/j.jhsg.2024.02.009.

[7] J. Rochotte, A. Sanap, V. Silenzio, and V. K. Singh, "Predicting anxiety using Google and Youtube digital traces," *Emerg. Trends Drugs, Addict. Heal.*, vol. 4, no. February, p. 100145, 2024, doi: 10.1016/j.etdah.2024.100145.

[8] M. Jamil Reza, "Persepsi Mahasiswa terhadap Penggunaan Youtube sebagai Media Konten Video Kreatif," *J. Komun. dan Organ. (J-KO*, vol. 3, pp. 39–46, 2021.

[9] H. Ulya, "KOMODIFIKASI PEKERJA PADA YOUTUBER PEMULA DAN UNDERRATED (Studi Kasus YouTube Indonesia)," *Interak. J. Ilmu Komun.*, vol. 8, no. 2, p. 1, 2019, doi: 10.14710/interaksi.8.2.1-12.

[10] R. E. Sakti, "Rancangan Perpres Jurnalisme Berkualitas dan Pengalaman Negara Lain," *kompas.id*, 2023. https://www.kompas.id/baca/riset/2023/08/19/rancangan-perpres-jurnalisme-berkualitas-dan-pengalaman-negara-lain

[11] L. A. Ziyad, "Kiamat Kreator Konten dibalik PerPres Jurnalisme Berkualitas.pdf," *ERA.id*, 2023.

[12] M. Browning, "Sebuah rancangan peraturan berpotensi mengancam masa depan media di Indonesia," *indonesia.googleblog*, 2023. https://indonesia.googleblog.com/2023/07/rancangan-peraturan-untuk-masa-depan-media-di-Indonesia.html

[13] T. K. Tran and T. T. Phan, "Capturing Contextual Factors in Sentiment Classification: An Ensemble Approach," *IEEE Access*, vol. 8, pp. 116856–116865, 2020, doi: 10.1109/ACCESS.2020.3004180.

[14] A. Mustofa and R. Novita, "KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP PEMBERLAKUAN PEMBATASAN KEGIATAN MASYARAKAT MENGGUNAKAN TEXT MINING PADA TWITTER," vol. 99, no. 99, pp. 1–10, 2022, doi: 10.47065/bits.v9i9.999.

[15] Z. Grotenhuis, P. J. Mosteiro, and A. M. Leeuwenberg, "Modest performance of text mining to extract health outcomes may be almost sufficient for high-quality prognostic model development," *Comput. Biol. Med.*, vol. 170, no. July 2023, p. 108014, 2024, doi: 10.1016/j.compbiomed.2024.108014.

[16] C. J. Varshney, A. Sharma, and D. P. Yadav, "Sentiment analysis using ensemble classification technique," *2020 IEEE Students' Conf. Eng. Syst. SCES 2020*, pp. 12–17, 2020, doi: 10.1109/SCES50439.2020.9236754.

[17] S. González, S. García, J. Del Ser, L. Rokach, and F. Herrera, "A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities," *Inf. Fusion*, vol. 64, no. July, pp. 205–237, 2020, doi: 10.1016/j.inffus.2020.07.007.

[18] I. D. Mienye, Y. Sun, and S. Member, "A Survey of Ensemble Learning : Concepts , Algorithms , Applications , and Prospects," *IEEE Access*, vol. 10, no. August, pp. 99129–99149, 2022, doi: 10.1109/ACCESS.2022.3207287.

[19] M. Kabir, M. M. J. Kabir, S. Xu, and B. Badhon, "An empirical research on sentiment analysis using machine learning approaches," *Int. J. Comput. Appl.*, vol. 43, no. 10, pp. 1011–1019, 2021, doi: 10.1080/1206212X.2019.1643584.

[20] A. Rahmadeyan, Mustakim, I. Ahmad, A. D. Alexander, and A. Rahman, "Phishing Website Detection with Ensemble Learning Approach Using Artificial Neural Network and AdaBoost," *2023 Int. Conf. Inf. Technol. Res. Innov. ICITRI 2023*, pp. 162–166, 2023, doi: 10.1109/ICITRI59340.2023.10249799.

[21] M. S. Md Suhaimin, M. H. Ahmad Hijazi, E. G. Moung, P. N. E. Nohuddin, S. Chua, and F. Coenen, "Social media sentiment analysis and opinion mining in public security: Taxonomy, trend analysis, issues and future directions," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 9, p. 101776, 2023, doi: 10.1016/j.jksuci.2023.101776.

[22] M. K. Iqbal, K. Abid, S. u din Ayubi, N. Aslam, and others, "Omicron Tweet Sentiment Analysis Using Ensemble Learning," *J. Comput. Biomed. Informatics*, vol. 4, no. 02, pp. 160–171, 2023.

[23] M. Qorib, T. Oladunni, M. Denis, E. Ososanya, and P. Cotae, "Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset," *Expert Syst. Appl.*, vol. 212, no. August 2022, p. 118715, 2023, doi: 10.1016/j.eswa.2022.118715.

[24] C. A. Agustina and R. Novita, "ScienceDirect The Implementation of TF-IDF and Word2Vec on Booster Vaccine Sentiment Analysis Using Support Vector Machine Algorithm," *Procedia Comput. Sci.*, vol. 234, pp. 156–163, 2024, doi: 10.1016/j.procs.2024.02.162.

[25] J. Kazmaier and J. H. van Vuuren, "The power of ensemble learning in sentiment analysis," *Expert Syst. Appl.*, vol. 187, no. June 2021, p. 115819, 2022, doi: 10.1016/j.eswa.2021.115819.

[26] V. Ahuja and M. Shakeel, "Twitter Presence of Jet Airways-Deriving Customer Insights Using Netnography and Wordclouds," *Procedia Comput. Sci.*, vol. 122, pp. 17–24, 2017, doi: 10.1016/j.procs.2017.11.336.

[27] F. Bima, "Exploratory Data Analysis of Indonesian Presidential Election Candidate Campaign in 2019 on Twitter," *Indones. J. Artif. Intell. Data Min.*, vol. 7, no. 2, pp. 229–240, 2024.