



Analisis Sentimen Persepsi Publik terhadap Tagar #KaburAjaDulu di Twitter Menggunakan Metode Lexicon-Based dengan Kamus InSet

S.M Zaidan Ali Difyah^{1*}, Eko Arip Winanto^{2*}, Pareza Alam Jusia³, Sharipuddin⁴

^{1,2,3}Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dinamika Bangsa, Jl. Jend. Sudirman, The Hok, Kec. Jambi Sel., Kota Jambi, Jambi 36138, Indonesia.

* Penulis Korespondensi, Email: winanto@unama.ac.id

Abstrak– Media sosial kerap menjadi cerminan respons publik terhadap isu sosial dan politik. Twitter, sebagai salah satu platform paling aktif, memfasilitasi percakapan singkat melalui penggunaan hashtag yang mencerminkan sentimen kolektif masyarakat. Penelitian ini bertujuan menganalisis persepsi publik terhadap tagar #KaburAjaDulu di Twitter menggunakan pendekatan lexicon-based dengan kamus InSet (Indonesia Sentiment Lexicon). Sebanyak 581 tweet berbahasa Indonesia yang memuat tagar tersebut dikumpulkan pada periode 1 Januari 2025 hingga 8 Mei 2025, kemudian diproses melalui tahapan pra-pemrosesan teks, yang mencakup case folding, pembersihan karakter khusus, tokenisasi, stopword removal (termasuk penghapusan kata “kaburajadulu”), serta lemmatisasi. Setiap tweet dianalisis melalui pencocokan token terhadap entri dalam kamus InSet untuk menghitung skor polaritas, yang kemudian diklasifikasikan ke dalam tiga kategori sentimen: positif, negatif, dan netral. Hasil klasifikasi menunjukkan distribusi sentimen dengan 41,72% negatif, 33,73% netral, dan 24,55% positif. Analisis terhadap kata-kata dominan mengindikasikan bahwa tagar tersebut banyak digunakan untuk mengekspresikan kritik, sarkasme, serta ketidakpuasan publik terhadap isu sosial dan politik nasional. Pendekatan lexicon-based terbukti efisien dalam menganalisis opini publik berbahasa Indonesia karena tidak memerlukan proses pelatihan model (training) dan mampu memberikan gambaran kuantitatif yang jelas terhadap kecenderungan sentimen masyarakat. Namun demikian, metode ini masih memiliki keterbatasan dalam mengenali konteks sarkasme dan negasi pada bahasa informal yang digunakan di Twitter. Secara keseluruhan, penelitian ini memberikan kontribusi empiris berupa pemetaan persepsi publik terhadap fenomena digital #KaburAjaDulu, sekaligus menjadi pijakan bagi pengembangan metode analisis sentimen yang lebih adaptif dan kontekstual pada studi komunikasi digital di masa mendatang.

Kata Kunci: Analisis Sentimen; Twitter; lexicon-Based; Kamus Inset; Google Colaboratory.

Abstract– Social media often reflects public responses to social and political issues. Twitter, as one of the most active platforms, facilitates short conversations through the use of hashtags that reflect the collective sentiment of society. This study aims to analyze public perceptions of the hashtag #KaburAjaDulu on Twitter using a lexicon-based approach with the InSet (Indonesia Sentiment Lexicon) dictionary. A total of 581 Indonesian-language tweets containing the hashtag were collected between January 1, 2025, and May 8, 2025, and then processed through text pre-processing stages, which included case folding, special character cleaning, tokenization, stopword removal (including removal of the word “kaburajadulu”), and lemmatization. Each tweet was analyzed by matching tokens to entries in the InSet dictionary to calculate a polarity score, which was then classified into three sentiment categories: positive, negative, and neutral. The classification results showed a sentiment distribution of 41.72% negative, 33.73% neutral, and 24.55% positive. Analysis of the dominant words indicates that the hashtag is widely used to express criticism, sarcasm, and public dissatisfaction with national social and political issues. The lexicon-based approach has proven to be efficient in analyzing Indonesian-language public opinion because it does not require model training and is able to provide a clear quantitative picture of public sentiment trends. However, this method still has limitations in recognizing the context of sarcasm and negation in the informal language used on Twitter. Overall, this study provides empirical contributions in the form of mapping public perceptions of the #KaburAjaDulu digital phenomenon, as well as a foundation for the development of more adaptive and contextual sentiment analysis methods in future digital communication studies.

Keywords: Sentimen Analysis; Twitter; Lexicon-Based; Inset Lexicon; Google Colaboratory.

1. PENDAHULUAN

Media sosial telah berkembang menjadi ruang publik digital yang tidak hanya berfungsi sebagai sarana komunikasi, tetapi juga sebagai wadah penyampaian opini masyarakat terhadap isu sosial, politik, dan ekonomi. Twitter, sebagai salah satu platform dengan pengguna aktif terbanyak di Indonesia, memungkinkan warganet untuk mengekspresikan pandangan mereka secara langsung melalui format teks singkat yang sering kali memuat kritik, sarkasme, maupun dukungan terhadap suatu isu [1]. Fenomena ini menjadi semakin menarik ketika muncul tagar tertentu yang kemudian viral, seperti tagar #KaburAjaDulu, yang mencerminkan respons publik terhadap kondisi sosial-politik nasional [2]. Analisis sentimen muncul sebagai salah satu metode yang relevan untuk

mengkaji opini publik di media sosial. Analisis ini berfungsi untuk mengelompokkan opini ke dalam kategori positif, negatif, atau netral, sehingga dapat memberikan gambaran umum tentang sikap kolektif masyarakat terhadap suatu fenomena [3]. Pendekatan ini sangat penting karena opini publik yang diekspresikan di media sosial sering kali berimplikasi terhadap dinamika sosial dan kebijakan publik [4].

Beberapa penelitian terdahulu telah menerapkan metode analisis sentimen di media sosial dengan berbagai pendekatan. Misalnya, penelitian yang dilakukan oleh Kusumaningsih dkk. menganalisis sentimen komentar di Twitter terkait pemilu dengan bantuan *tool* MonkeyLearn, dan hasilnya menunjukkan kemampuan analisis sentimen untuk memetakan kecenderungan opini politik masyarakat [5]. Penelitian lain oleh Larasati dkk. meneliti ulasan aplikasi Dana menggunakan metode *Random Forest*, yang meskipun akurat namun memerlukan data latih dalam jumlah besar [6]. Penelitian berbasis *lexicon-based* juga banyak dilakukan. Bagas dkk. menerapkan metode ini untuk menganalisis sentimen pengguna Twitter terhadap kinerja ISP di Indonesia, dan hasilnya menunjukkan bahwa metode *lexicon-based* dapat digunakan secara praktis tanpa memerlukan proses pelatihan [7]. Hal serupa dilakukan oleh Ratnaswari dkk., yang mengombinasikan metode *lexicon-based* dengan *Support Vector Machine* untuk menganalisis opini publik terkait Presiden dan Wakil Presiden periode 2024–2029 [8].

Lebih lanjut, Hendrawati dkk. menggunakan metode *lexicon-based* untuk menelaah sentimen publik terkait larangan impor pakaian bekas, yang menunjukkan bahwa isu sosial-ekonomi juga dapat dipetakan menggunakan pendekatan ini [9]. Sementara itu, Aulia dkk. meneliti tagar *#IndonesiaTerseher* pada masa pandemi COVID-19 dengan metode SentiStrength, yang mengungkap pola dominasi sentimen negatif masyarakat [2].

Penelitian lain yang relevan adalah karya Syakur, yang menggunakan metode *lexicon-based* untuk menganalisis respons publik terhadap kebijakan pemerintah dalam pencegahan penyebaran COVID-19 [4]. Zahria dkk. juga melakukan studi tentang Twitter sebagai platform komunikasi politik di Indonesia, menegaskan bahwa media sosial berperan penting dalam pembentukan opini publik terhadap isu kebangsaan [1].

Selain itu, Hernikawati meneliti respons masyarakat terhadap vaksin Sinovac melalui *lexicon-based sentiment analysis*, yang menekankan bagaimana opini publik terhadap kebijakan kesehatan dapat dipetakan secara sistematis [10]. Amaliah dkk. membandingkan akurasi metode *lexicon-based* dan *Naïve Bayes Classifier* pada analisis opini publik terhadap aplikasi investasi di Twitter, dan menemukan bahwa *lexicon-based* lebih sederhana dalam implementasi [11]. Relevansi kamus sentimen untuk Bahasa Indonesia juga dibuktikan oleh Koto dan Rahmaningtyas melalui pengembangan *InSet Lexicon*, yang secara khusus dirancang untuk menangani analisis sentimen pada teks berbahasa Indonesia [12]. Evaluasi ini menegaskan bahwa *InSet* dapat menjadi solusi praktis dalam penelitian yang tidak membutuhkan data latih yang besar, namun tetap mampu menangkap polaritas teks di media sosial.

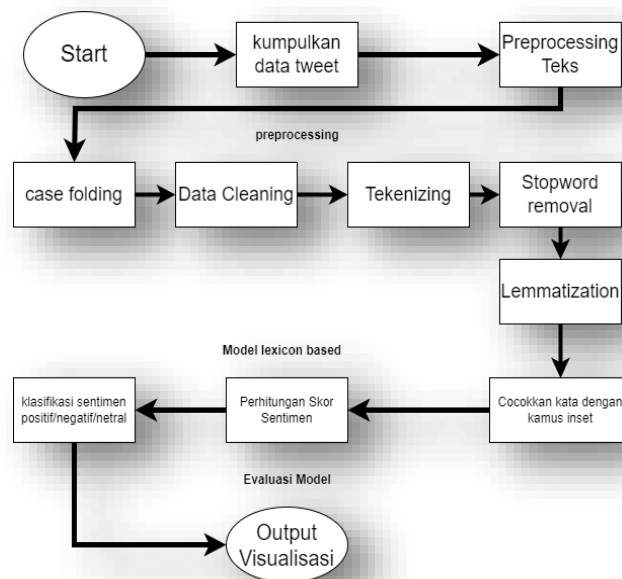
Penelitian Yuliasuti dkk. turut menekankan penerapan *InSet Lexicon* pada analisis kenaikan harga BBM, yang memperlihatkan efektivitas pendekatan ini dalam menelaah isu ekonomi di ruang digital [13]. Firdaus dkk. juga menegaskan manfaat pendekatan *lexicon-based* dalam evaluasi opini mahasiswa, yang memperlihatkan bahwa metode ini dapat digunakan pada berbagai konteks [14]. Selain itu, Fauziah dkk. dalam *systematic literature review* menegaskan bahwa *lexicon-based sentiment analysis* tetap relevan digunakan dalam konteks Bahasa Indonesia meskipun memiliki keterbatasan dalam mendeteksi sarkasme [15]. Hal ini sejalan dengan penelitian Prasetya dkk. yang menyoroti penerapan *lexicon-based* untuk isu COVID-19 di Twitter, yang memperlihatkan bahwa meskipun efisien, metode ini membutuhkan interpretasi lebih lanjut untuk menangkap nuansa komunikasi digital masyarakat [16]. Dari berbagai penelitian terdahulu tersebut dapat dilihat bahwa pendekatan *lexicon-based* telah banyak digunakan pada isu politik, ekonomi, hingga kesehatan, namun penelitian yang secara khusus menelaah fenomena tagar *#KaburAjaDulu* masih terbatas. Padahal, tagar ini memiliki nilai penting karena mencerminkan ekspresi publik yang khas dengan nuansa kritik dan sarkasme. Celah inilah yang menjadi *gap analysis* penelitian ini, sehingga diperlukan kajian yang lebih fokus menggunakan *InSet Lexicon* untuk memahami bagaimana masyarakat digital Indonesia menanggapi isu sosial-politik melalui tagar tersebut. Penelitian ini tidak hanya menerapkan metode *lexicon-based* pada isu baru, tetapi juga meninjau karakteristik ekspresi publik khas Indonesia seperti sarkasme dan ironi dalam teks media sosial, yang belum banyak dikaji sebelumnya.

Berdasarkan latar belakang dan tinjauan penelitian terdahulu, tujuan dari penelitian ini adalah untuk mengklasifikasikan sentimen publik terhadap tagar *#KaburAjaDulu* menggunakan pendekatan *lexicon-based* berbasis kamus *InSet*. terlihat bahwa metode *lexicon-based* dan kamus *InSet* terbukti efektif untuk teks berbahasa Indonesia, terutama dalam konteks ulasan produk dan opini politik. Penelitian-penelitian tersebut menjadi landasan pemilihan metode pada studi ini, sekaligus membuka ruang untuk eksplorasi lanjutan dalam isu sosial digital seperti fenomena tagar *#KaburAjaDulu*. Kontribusi keilmuan dari penelitian ini terletak pada penerapan metode *lexicon-based* berbasis kamus *InSet* untuk menganalisis wacana sosial-politik yang berkembang di media sosial Indonesia. Selain memperluas cakupan penggunaan *InSet* dari ranah evaluasi produk menuju kajian sosial-politik, penelitian ini juga menambahkan dimensi analisis terhadap pola sarkasme dan kritik implisit dalam bahasa informal masyarakat Indonesia di Twitter. Dengan demikian, penelitian ini tidak hanya menerapkan metode yang sudah ada, tetapi juga memperluas pendekatan *lexicon-based* untuk konteks kebahasaan dan sosial yang lebih kompleks.

2. METODOLOGI PENELITIAN

Penelitian ini merujuk pada rangkaian langkah terstruktur yang dirancang untuk mengevaluasi sentimen publik terhadap tagar #KaburAjaDulu di Twitter dengan pendekatan *lexicon-based*. Proses diawali dengan pengumpulan data berupa tweet yang mengandung tagar tersebut, kemudian dilanjutkan ke tahap pra-pemrosesan yang berfungsi untuk menyaring, membersihkan, dan menyiapkan data agar dapat dianalisis secara optimal. Setelah data siap, setiap kata dalam tweet akan dicocokkan dengan entri dalam kamus InSet (Indonesia Sentiment Lexicon) guna memperoleh nilai polaritas. Hasil pencocokan ini akan digunakan untuk menghitung total skor sentimen pada setiap tweet. Berdasarkan skor tersebut, tweet kemudian diklasifikasikan ke dalam kategori sentimen tertentu, seperti positif, negatif, atau netral. Seluruh hasil klasifikasi ini selanjutnya divisualisasikan dalam bentuk grafik atau tabel, yang akan menjadi dasar interpretasi dalam analisis akhir.

Penjabaran teknis mengenai setiap tahapan dalam alur ini, termasuk representasi visual berupa flowchart serta pseudocode, akan dibahas lebih lanjut pada bagian-bagian berikutnya dalam dokumen penelitian ini. Adapun alur eksperimen yang digunakan dalam penelitian ini dapat dilihat pada Gambar 1:



Gambar 1. Tahapan Metode Penelitian

2.1 Pengumpulan Dataset

Proses pengumpulan data dalam penelitian ini dilakukan dengan memanfaatkan teknik web scraping melalui bantuan pustaka Python, salah satunya *tweet harvest*, yang memungkinkan penelusuran dan ekstraksi data dari platform media sosial Twitter. Fokus utama pengambilan data diarahkan pada tweet-tweet yang menyertakan tagar #KaburAjaDulu, karena dianggap mencerminkan berbagai bentuk pandangan masyarakat terhadap isu-isu sosial, politik, maupun ekonomi yang sedang berkembang. Pengambilan data dilakukan pada rentang waktu 1 Januari 2025, dengan hasil sebanyak 581 tweet berbahasa Indonesia yang berhasil dihimpun.

2.2 Preprocessing

Pra-pemrosesan data merupakan langkah awal yang sangat krusial dalam proses analisis sentimen, khususnya ketika menggunakan pendekatan berbasis leksikon. Tahapan ini bertujuan untuk menyiapkan teks agar dapat diproses lebih lanjut oleh sistem, terutama dalam hal pencocokan kata dengan kamus sentimen seperti InSet. Setiap tweet yang berhasil dikumpulkan akan dibersihkan dari unsur-unsur yang tidak diperlukan dan disesuaikan bentuknya agar lebih sederhana dan mudah dianalisis. Hal serupa juga disampaikan oleh Abdillah et al. (2021) dalam studi mereka mengenai persepsi publik terhadap kebijakan penanganan COVID-19 di Indonesia. Penelitian tersebut menggunakan pendekatan *lexicon-based* dan menemukan bahwa keberhasilan analisis sangat ditentukan oleh kualitas tahapan pra-pemrosesan, khususnya dalam menangani bentuk-bentuk kata yang tidak baku[17]. Berikut penjelasan tahapan yang dilakukan dalam *preprocessing* :

a. Case Folding

- Semua karakter huruf pada teks tweet diubah menjadi huruf kecil (*lowercase*). Tujuannya adalah untuk menyeragamkan kata-kata, sehingga tidak ada perbedaan makna antara kata yang ditulis dengan huruf kapital maupun huruf kecil, seperti pada kata “Negara” dan “negara” yang akan dianggap sama.
- b. *Pembersihan Teks (Cleaning)*
Pada tahap ini, elemen-elemen yang bukan bagian dari isi utama teks, seperti tautan (URL), *username* (@), tagar (#), angka, tanda baca, dan simbol-simbol lainnya dihapus. Hal ini dilakukan agar teks menjadi lebih bersih dan hanya menyisakan kata-kata inti yang mungkin mengandung muatan sentimen.
 - c. *Tokenisasi (Tokenizing)*
Kalimat dalam tweet kemudian dipecah menjadi satuan kata-kata terpisah. Proses ini penting untuk memungkinkan setiap kata dianalisis secara individual, sehingga bisa dicocokkan dengan entri yang ada dalam kamus InSet.
 - d. *Stopword Removal*
Kata-kata umum dalam bahasa Indonesia yang tidak memiliki kontribusi terhadap nilai sentimen seperti “yang”, “di”, “dengan”, “itu”, dan sejenisnya akan dihapus dari teks. Penghilangan *stopword* bertujuan untuk menyaring informasi yang tidak memiliki nilai emosional, agar proses analisis menjadi lebih fokus dan efisien.
 - e. *Lemmatization/stemming*
Kata-kata diubah ke dalam bentuk dasarnya (akar kata). Misalnya, “berlari”, “pelari”, dan “berlarian” akan disederhanakan menjadi “lari”. Proses ini penting untuk menyelaraskan bentuk kata dalam teks dengan entri yang tersedia di dalam kamus InSet. Tabel 1 menyajikan hasil *preprocessing*.

Tabel 1. Contoh Hasil *Preprocessing*

DATA SEBELUM <i>PREPROCESSING</i>	DATA SETELAH <i>PREPROCESSING</i>
i wish 2 tahunnn kemudian bisa ikut trend #kaburajadulu	i wish tahunnn kemudian ikut trend
kpn bisa join trend #kaburajadulu	kpn join trend
@RWWReborn @Blackburger_28 Peserta #kaburajadulu dari kalangan ijo kah?	peserta kalangan ijo kah
bayi memilih #KaburAjaDulu daripada terlahir jadi WNI	bayi memilih daripada terlahir jadi wni

2.3 Model lexicon-based

Pendekatan *lexicon-based* dalam analisis sentimen merupakan metode yang bergantung pada daftar kata atau kamus yang telah diberi label polaritas, baik berupa sentimen positif, negatif, maupun netral. Pada penelitian ini, alat bantu yang digunakan berupa kamus InSet (Indonesia Sentiment Lexicon), yaitu sebuah sumber leksikal berbahasa Indonesia yang bersifat terbuka dan telah dikembangkan secara publik. InSet disusun oleh Fajri Koto dan Gemala Y. Rahmaningtyas pada tahun 2017, kemudian dipublikasikan melalui 21st International Conference on Asian Language Processing (IALP). Kamus ini memuat lebih dari 3.600 kata positif dan sekitar 6.600 kata negatif, masing-masing memiliki bobot skor polaritas dalam rentang -5 hingga +5. Repositori lengkap InSet tersedia secara terbuka melalui GitHub ([fajri91/InSet](https://github.com/fajri91/InSet)), sehingga dapat diakses dan dimanfaatkan secara luas oleh peneliti maupun praktisi. Kamus ini memuat ratusan entri kosakata yang masing-masing dilengkapi dengan nilai sentimen yang mencerminkan muatan emosional dari setiap kata tersebut.

Sejumlah kajian terdahulu [15] menyebutkan bahwa InSet merupakan salah satu leksikon sentimen yang paling banyak digunakan dalam analisis berbahasa Indonesia karena menawarkan akurasi tinggi serta kemudahan penerapan. Adapun hasil riset lain [18] menunjukkan bahwa penerapan InSet dalam analisis sentimen terhadap pembelajaran daring mampu menghasilkan klasifikasi sentimen yang akurat, bahkan dalam konteks bahasa yang sangat informal. Hal serupa disampaikan pula pada temuan penelitian [14], yang menyimpulkan bahwa InSet memberikan hasil lebih unggul dibandingkan leksikon lain seperti SentiWordNet versi terjemahan, khususnya dalam analisis umpan balik mahasiswa terhadap pengajaran dosen.

Sementara itu, studi lain [19] yang mengembangkan aplikasi web berbasis *lexicon* untuk analisis sentimen dalam Bahasa Indonesia juga memilih InSet sebagai leksikon utama karena kelengkapan kosakata dan kesesuaiannya dengan bahasa sehari-hari yang digunakan di media sosial.

2.4 Evaluasi Hasil Analisis Sentimen

Tahap ini tidak melibatkan proses pelatihan model (training), melainkan evaluasi terhadap hasil klasifikasi sentimen yang diperoleh melalui pendekatan *lexicon-based*. Evaluasi dilakukan dengan meninjau distribusi sentimen (positif, negatif, dan netral), serta menginterpretasikan makna di balik kata-kata dominan yang muncul. Analisis ini bertujuan untuk melihat kecenderungan opini publik terhadap isu sosial yang direpresentasikan melalui tagar #KaburAjaDulu.

3. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan analisis sentimen publik terhadap tagar #KaburAjaDulu di Twitter dengan memanfaatkan pendekatan *lexicon-based* menggunakan kamus InSet (Indonesia Sentiment Lexicon). Proses analisis dilakukan melalui serangkaian tahapan, meliputi pengumpulan data tweet dengan *Tweet Harvest*, pra-pemrosesan teks (pembersihan, *normalisasi*, *stopword removal*, dan *lemmatisasi*), hingga klasifikasi sentimen ke dalam kategori positif, negatif, dan netral. Implementasi metode ini dilakukan menggunakan *Google Colaboratory* serta pustaka *Python* untuk mendukung proses otomatisasi analisis dan visualisasi hasil.

3.1 Deskripsi Dataset

Pengumpulan data dilakukan dalam rentang waktu mulai dari 1 Januari 2025, dan dari proses tersebut berhasil dihimpun sebanyak 581 tweet yang ditulis dalam Bahasa Indonesia. Setiap entri dalam dataset tersebut memuat informasi berupa teks asli dari cuitan atau *full_text*, yang menjadi objek Sebagai gambaran awal, pada tabel 2, disajikan cuplikan data sebagaimana bentuk aslinya sebelum melalui proses preprocessing.

Tabel 2. Contoh Dataset

No	Sebelum
1	Marinir pun bisa #kaburAjaDulu gaji lebih tinggi tunjangan lebih baik peralatan lebih canggih Minus nya apa?
2	Tim yang menggaungkan #kaburajadulu mana nih suaranya? Diem diem bae apa malah fokus ke teori konspirasi?
3	Ini kan kalian yang berisik pengen #kaburajadulu FYI aja kalo kalian pernah kena tbc kalian gakn diterima di negara manapun mau sekolah/kerja dll Ini pemrentah mau penelitian biar ada vaksinnnya kalian protes juga klean emang pantes hidup susah karena kegoblogan klean sendiri
4	yang mau kuliah ke china S1/S2 bisa dm aku yaaa. Open pendaftaran sampe akhir bulan may nanti dibantu cari beasiswa (tapi tergantung sama nilai juga). #kaburajadulu #beasiswachina #sekolahkechina #beasiswachina2025 #utbk2025!
5	Jadi pengen ngikut tren #kaburajadulu wkwkwkwk Selain ga betah sama birokrasi yg awut-awutan aku juga ga betah sama budaya yg kayak gini

3.2 Preprocessing

Dataset yang sebelum nya berjumlah 581 mengalami penyusutan karna telah melalui tahap preprocessing (*Case Folding*, *Data Cleaning*, *Tokenisasi*, *Stopword Removal*, *Lemmatization/Stemming*). Pada tahap *stopword removal* merupakan impact yang sangat besar terhadap *preprocessing*. Pada tabel 3 disajikan beberapa kata yang ada di *stopwrod removal*:

Tabel 3. Kata-kata Stopword Removal

No	Kelompok Kata	Contoh Kata Tambahan	Keterangan
1	Kata ganti informal	gw, gua, lu, kalian, saya, kamu, kami, kita, dia, mereka	Bentuk kata ganti nonbaku atau umum dalam percakapan Twitter yang tidak memiliki makna sentimen.
2	Pertikel informasi	kan, deh, dong, tuh, lah, aja, nih, sih, loh	Partikel penegas dalam bahasa informal yang tidak berkontribusi pada polaritas emosi.

3	Kata hubung umum	jadi, buat, tapi, agar, maka, dan, atau, untuk, karena, juga, hanya, lebih, lagi	Kata sambung atau penghubung yang berfungsi gramatikal tanpa membawa makna emosional.
4	Kata tanya dan fungsional	apa, siapa, kenapa, gimana, yg, ada, di, ke, dari, dalam, pada, akan, bisa, sudah, tidak, ya, itu, ini, mau	Kata interogatif, preposisi, atau kata bantu yang berfungsi struktural dan tidak mengandung opini.
5	Ekpresi non leksikal	wkwk, haha, hehe, yah, oh, eh, oke, pls	Ekpresi atau interjeksi spontan yang tidak mencerminkan emosi positif atau negatif dalam konteks analisis.
6	Variasi penulisan informal	ajaa, yahh, kaburajadulu	Bentuk ejaan tidak baku atau kata tagar yang harus dihapus agar hasil sentimen tidak bias terhadap topik utama.

jumlah tweet berkurang menjadi 501 tweet, penyusutan ini terjadi karena adanya data yang duplikat, tweet kosong, maupun teks yang kehilangan makna. Seluruh tweet disimpan dalam format Comma Separated Values (CSV), yang memungkinkan pengolahan data lebih mudah pada tahap selanjutnya. Tabel 4 menyajikan hasil perubahan dataset yang telah melalui proses *preprocessing*.

Tabel 4. Hasil *Preprocessing*

No	Sebelum	Sesudah	Tokens
1	Marinir pun bisa #kaburAjaDulu gaji lebih tinggi tunjangan lebih baik peralatan lebih canggih Minus nya apa?	marinir pun gaji tinggi tunjangan baik peralatan canggih minus nya apa	['marinir', 'pun', 'gaji', 'tinggi', 'tunjangan', 'baik', 'peralatan', 'canggih', 'minus', 'nya', 'apa']
2	Tim yang menggaungkan #kaburajadulu mana nih suaranya? Diem diem bae apa malah fokus ke teori konspirasi?	tim menggaungkan mana nih suaranya diem bae apa malah fokus teori konspirasi	['tim', 'menggaungkan', 'mana', 'nih', 'suaranya', 'diem', 'bae', 'apa', 'malah', 'fokus', 'teori', 'konspirasi']
3	Ini kan kalian yang berisik pengen #kaburajadulu FYI aja kalo kalian pernah kena tbc kalian gak diterima di negara manapun mau sekolah/kerja dll Ini pemerintah mau penelitian biar ada vaksinnya kalian protes juga klean emang pantes hidup susah karena kegoblogan klean sendiri	kan kalian berisik pengen fyi kalo pernah kena tbc gakan diterima negara manapun mau sekolahkerja dll pemerintah penelitian biar ada vaksinnya protes klean emang pantes hidup susah kegoblogan sendiri	['kan', 'kalian', 'berisik', 'pengen', 'fyi', 'kalo', 'pernah', 'kena', 'tbc', 'gakan', 'diterima', 'negara', 'manapun', 'mau', 'sekolahkerja', 'dll', 'pemerintah', 'penelitian', 'biar', 'ada', 'vaksinnya', 'protes', 'klean', 'emang', 'pantes', 'hidup', 'susah', 'kegoblogan', 'sendiri']
4	yang mau kuliah ke china S1/S2 bisa dm aku yaaa. Open pendaftaran sampe akhir bulan may nanti dibantu cari beasiswa (tapi tergantung sama nilai juga). #kaburajadulu #beasiswachina #sekolahkechina #beasiswachina2025 #utbk2025!	mau kuliah china ss dm aku yaaa open pendaftaran sampe akhir bulan may nanti dibantu cari beasiswa tapi tergantung sama nilai beasiswachina sekolahkechina utbk	['mau', 'kuliah', 'china', 'ss', 'dm', 'aku', 'yaaa', 'open', 'pendaftaran', 'sampe', 'akhir', 'bulan', 'may', 'nanti', 'dibantu', 'cari', 'beasiswa', 'tapi', 'tergantung', 'sama', 'nilai', 'beasiswachina', 'sekolahkechina', 'utbk']
5	Jadi pengen ngikut tren #kaburajadulu wkwkwkwk Selain ga betah sama birokrasi yg awut-awutan aku juga ga betah sama budaya yg kayak gini	jadi pengen ngikut tren wkwkwkwk selain ga betah sama birokrasi yg awutawutan aku budaya kayak gini	['jadi', 'pengen', 'ngikut', 'tren', 'wkwkwkwk', 'selain', 'ga', 'betah', 'sama', 'birokrasi', 'yg', 'awutawutan', 'aku', 'budaya', 'kayak', 'gini']

3.3 Hasil Metode Lexicon-Based

Berikut hasil yang diperoleh dari metode lexicon-based, dan beberapa hasil nya yaitu sebagai berikut :

3.3.1 Pencocokan kata dengan kamus inset

Setelah dataset yang telah melewati tahapan preprocessing akan di proses lagi menjadi kata kata tersendiri. Setiap kata yang akan di cocokkan dengan dataset yang terdapat di dalam kamus inset. Bila kata yang di cocokkan ada dalam kamus, maka akan di beri nilai sesuai bobot nya, seperti pada tabel 5.

Tabel 5. Contoh Kamus Sentimen

NO	KATA KATA NEGATIF	BOBOT	KATA KATA POSITIF	BOBOT
1	Tersentuh (perasaan)	-1	Paripurna	1
2	Putus tali gantung	-2	Merekam	2
3	Larat hati	-3	Hai	3
4	Linu	-4	Welas	4
5	Remuk redam	-5	terimakasih	5

3.3.2 Perhitungan Skor Sentimen

Setelah setiap kata di berikan bobot berdasarkan kata kata yang terdapat di dalam kamus inset, maka tahap selanjut nya hasil kata kata pembobotan yang teradapat di dalam kalimat akan melewati tahapan perhitungan, dengan rumus sederhananya pada persamaan (1).

$$SKOR SENTIMEN = (JUMLAH KATA POSITIF) - (JUMLAH KATA NEGATIF) \quad (1)$$

Berikut contoh hasil perhitungan dari hasil preprocessing, dan setelah melakukan penerapan rumus yang terdapat dalam tahapan perhitungan skor sentimen :

```
Tweet yang dianalisis: tim menggaungkan kaburajadulu mana nih suaranya diem bae apa malah fokus teori konspirasi
Kata positif ditemukan: ['diem', 'fokus', 'konspirasi']
Kata negatif ditemukan: ['tim', 'diem', 'apa', 'fokus', 'konspirasi']
Skor sentimen: -2
Kategori Sentimen: Netral
```

Gambar 2. Hasil Perhitungan Sentimen

Dalam penelitian ini, perhitungan skor sentimen dilakukan dengan menghitung selisih jumlah kata positif dan negatif berdasarkan pencocokan dengan kamus InSet. Namun, pendekatan ini memiliki keterbatasan dalam menangani konteks negasi seperti kata ‘tidak’ atau ‘bukan’. Misalnya, kalimat ‘tidak senang’ tetap terhitung positif karena kata ‘senang’ terdapat dalam daftar leksikon positif. Kelemahan ini menjadi salah satu faktor yang dapat memengaruhi akurasi hasil klasifikasi, dan diakui sebagai ruang pengembangan untuk penelitian lanjutan agar dapat menangkap konteks negasi dan sarkasme secara lebih akurat.

3.3.3 Klasifikasi Sentimen

Skor sentimen yang di peroleh dari masing masing tweet kemudian di pakai untuk menentukan sentimen. Aturan klasifikasinya disajikan pada tabel 6.

Tabel 6. Klasifikasi Sentimen

NO	SKOR SENTIMEN	KATEGORI SENTIMEN
1	+1 hingga +5	Positif

2	0	Netral
3	-1 hingga -5	Negatif

Dan berikut hasil penerapan dari klasifikasi sentimen yang telah di lakukan pada hasil preprocessing :

Tabel 7. Hasil Klasifikasi negative

No	Cleaned_text	Score	Sentiment
1	yaa penting laaahh	0	netral
2	ngga kerja luar negeri tanpa uang peser pun	-1	negatif
3	tim menggaungkan mana nih	-2	negatif
4	suaranya diem bae apa malah fokus teori konspirasi konoha gelap also lbp ngapain kabur ln kalian jdi nd citizen disana mending disinilah kau kesimpulannya mrintah pindah giliran mau eh dilarang rezim aneh ada resign dgn gagahnya balik kandang lgi	-3	negatif
5	marinir pun gaji tinggi tunjangan baik peralatan canggih minus nya apa	-4	negatif
6	org baratny jg gamau cari pasangan pny anak bnyk yg lbh milih adopsi hewan nuclear family turun tingkat penduduk menurun hasilny politikny lg terguncang awkowkowkowkow hal u sbt bener sih wong nz	-5	negatif
7	jangan lupa pake masker polusi udara jakarta baik saja zona kuning sampe terkena penyakit asma	-6	negatif
8	amarah gw k pemerintah dgn brbagai macam brita brusaha gak ungkapin mncoba jaga ketikan takut hujat diaspora tau diri ikut nimbrung buat bantu info tmen trmasuk puncak kemarahan sbnerny minimal ada manfaatny	-7	negatif
9	cringe karakternya gegabah ingin menjatuhkan rezim menjadi martir sebagai korban penghilangan paksa padahal dua decade kemudian reformasinya dikorupsi generasi mudanya memilih daripada harus repot melawan yolo	-9	negatif

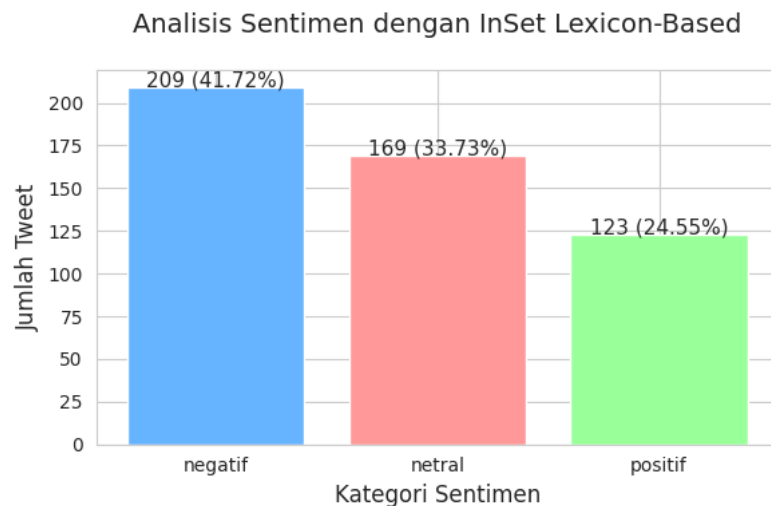
Tabel 8. Hasil Klasifikasi Positif

No	Cleaned_text	Score	Sentiment
1	ursa baby clear namanya tapi outputnya dewasa design transparan biar lo ngeliat isi eh podnya simple clean and lowkey stylish buat suka vibe slow burn tetap on point ursababyclear podkit vapestyle peringatandarurat	7	positif
2	terima kasih ganya kak aisyle amp day doa terbaik kalian berdua semoga akak ais segera mendapat pekerjaan proper aku menyusul aminn	5	positif
3	hi aku ada update kabar baik jadi kerjasama agensi pencarian kerja perawat khusus menyalurkan tenaga kesehatan untk feenya jt sdh dapat les bahasa online ab biaya proses penyetaraan translate dokumen tiket pesawat	4	positif
4	tindakan benar hasil penelitian negara indonesia kaya namun warganya miskin bahagia kaburajadulu studi harvard ri jadi warga paling makmur	3	positif
5	beruntung banget anda bu dok	2	positif
6	tuhan permudahkanlah amin	1	positif

3.4 Evaluasi Hasil Analisis Sentimen

Pada bagian ini disajikan Evaluasi hasil analisis sentimen #KaburAjaDulu di Twitter . Evaluasi dilakukan untuk melihat kemampuan model dalam mengelompokkan sentimen sebagai negatif, netral dan positif.

3.4.1 Visualisasi Sentimen

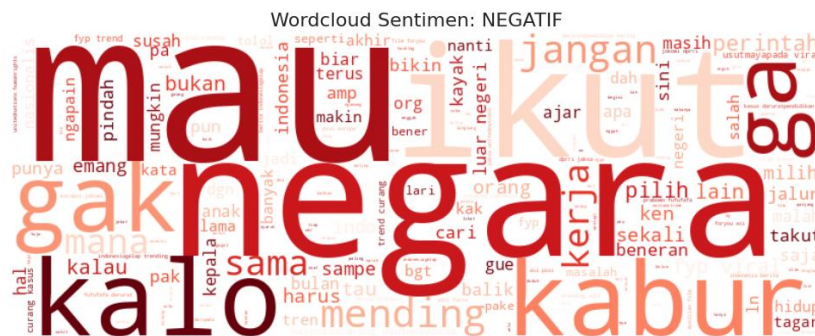


Gambar 3. Visualisasi klasifikasi sentimen (Diagram Batang)

Berdasarkan gambar 3 hasil klasifikasi sentimen terhadap 581 tweet yang mengandung tagar #KaburAjaDulu, diperoleh bahwa kategori negatif mendominasi sebesar 41,72%, disusul oleh sentimen netral sebesar 33,73%, dan positif sebesar 24,55%. Persentase ini menunjukkan bahwa tagar tersebut lebih banyak digunakan untuk mengekspresikan kritik, kekecewaan, atau sindiran terhadap situasi sosial dan politik. Sementara itu, meskipun sebagian tweet dikategorikan netral secara tekstual, banyak di antaranya mengandung makna sarkastik atau kritik terselubung yang tidak sepenuhnya terbaca oleh pendekatan leksikal. Hal ini mempertegas bahwa metode berbasis leksikon mampu memberikan klasifikasi kategorikal yang jelas, namun interpretasi hasil tetap perlu mempertimbangkan konteks sosial serta gaya bahasa informal yang umum digunakan masyarakat Indonesia di media sosial.

3.4.2 Analisis Kata Dominan

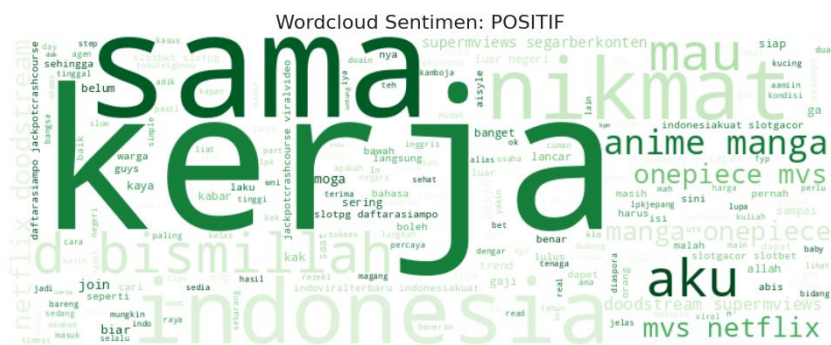
Agar mengetahui seperti apa saja kata kata yang teradapat di dalam sentimen negatif, netral, dan positif, maka disajikan bentuk visualisasi *wordcloud*, dari hasil sentimen yang di dapatkan setelah melewati berbagai tahapan yang telah di jalankan. Pada pada gambar 4 disajikan visualisasi sentimen negatif, pada gambar 5 untuk sentimen netral dan gambar 6 visualisasi sentimen positif.



Gambar 4. Visualisasi Wordcloud Negatif



Gambar 5. Visualisasi Wordcloud Netral



Gambar 6. Visualisasi Wordcloud Positif

Berdasarkan hasil wordcloud pada setiap kategori sentimen, ditemukan bahwa tagar #KaburAjaDulu memiliki kecenderungan kuat sebagai sarana ekspresi opini publik yang bersifat kritis. Pada kategori negatif, kata seperti “negara,” “kabur,” “mending,” dan “gak” sering muncul bersamaan, mencerminkan bentuk kekecewaan, ketidakpuasan, dan sarkasme terhadap isu sosial serta politik nasional. Sementara itu, kategori positif menampilkan kata seperti “kerja,” “bismillah,” dan “nikmat” yang cenderung menggambarkan optimisme personal atau semangat individu. Adapun kategori netral menampilkan kata seperti “aku,” “indonesia,” dan “orang,” yang cenderung deskriptif tanpa emosi yang kuat.

Temuan ini memperkuat interpretasi bahwa persepsi publik terhadap tagar #KaburAjaDulu lebih dominan digunakan untuk menyuarakan kritik sosial dan refleksi terhadap situasi nasional secara tidak langsung melalui gaya bahasa informal di Twitter.

4. KESIMPULAN

Berdasarkan hasil analisis terhadap 581 tweet dengan tagar #KaburAjaDulu, diperoleh distribusi sentimen sebesar 41,72% negatif, 33,73% netral, dan 24,55% positif. yang ditandai dengan kemunculan kata seperti “negara,” “kabur,” dan “mending.” Pola tersebut menggambarkan penggunaan tagar ini sebagai bentuk ekspresi kritik dan ketidakpuasan publik terhadap dinamika sosial dan politik di Indonesia. Sementara itu, sentimen positif dan netral lebih merepresentasikan respons personal dan ekspresi deskriptif tanpa muatan emosi yang kuat. Secara keseluruhan, hasil ini menegaskan bahwa metode lexicon-based efektif dalam mengungkap persepsi publik terhadap fenomena digital berbahasa Indonesia, meskipun masih memiliki keterbatasan dalam mengenali konteks sarkasme dan negasi.

REFERENCES

- [1] F. Zahria Emeraldien, R. Jefri Sunarsono, R. Alit, J. Raya Rungkut Madya, G. Anyar, and J. Timur, “TWITTER SEBAGAI PLATFORM KOMUNIKASI POLITIK DI INDONESIA,” jakarta, Feb. 2019. doi: <https://doi.org/10.33005/scan.v14i1.1457>.
- [2] B. Aulia, P. E. P. Utomo, U. Khaira, and T. Suratno, “ANALISIS SENTIMEN TAGAR #INDONESIATERSERAH DI MASA COVID-19 MENGGUNAKAN METODE SENTISTRENGTH,” *Jurnal Komputer dan Informatika*, vol. 9, no. 2, pp. 207–213, Oct. 2021, doi: 10.35508/jicon.v9i2.4275.
- [3] R. Arief and K. Imanuel, “ANALISIS SENTIMEN TOPIK VIRAL DESA PENARI PADA MEDIA SOSIAL TWITTER DENGAN METODE LEXICON BASED Universitas Gunadarma 1, 2 Jalan Margonda Raya No 100 Depok Jawa Barat 16424 Sur-el : rifiana@staff.gunadarma.ac.id 1 , karel4404@gmail.com 2,” *Jurnal Ilmiah MATRIK*, vol. 21, no. 3, 2019.

- [4] A. Syakur, "IMPLEMENTASI METODE LEXICON BASE UNTUK ANALISIS SENTIMEN KEBIJAKAN PEMERINTAH DALAM PENCEGAHAN PENYEBARAN VIRUS CORONA COVID-19 PADA TWITTER," *Jurnal Ilmiah Informatika Komputer*, vol. 26, no. 3, pp. 247–260, 2021, doi: 10.35760/ik.2021.v26i3.4720.
- [5] D. Kusumaningsih, H. G. Andaresta, K. Saddhono, and H. Hanafi, "Analisis Sentimen Komentar Twitter pada Pemilu," *Indonesia Berbantuan Situs Monkeylearn*, vol. 12, no. 2, pp. 145–156, 2024, doi: 10.25299/geram.2024.21484.
- [6] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," 2022. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [7] M. Bagas, D. Putra, and E. Setiawan, "METODE LEXICON BASED UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP KINERJA ISP (STUDI KASUS : INDIHOME, BIZNET, MYREPUBLIC)," 2024. [Online]. Available: <http://dev.twitter.com>
- [8] S. Ratnaswari, N. C. Wibowo, and D. S. Y. Kartika, "ANALISIS SENTIMEN MENGGUNAKAN METODE LEXICON-BASED DAN SUPPORT VECTOR MACHINE PADA PRESIDEN DAN WAKIL PRESIDEN INDONESIA PERIODE 2024–2029," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5604.
- [9] T. Hendrawati, N. L. W. S. R. Ginantra, and C. M. Saiman, "Analisis Sentimen Larangan Impor Pakaian Bekas Menggunakan Metode Support Vectore Machine dan Lexicon Based," *TEMATIK*, vol. 11, no. 1, pp. 56–64, Jun. 2024, doi: 10.38204/tematik.v11i1.1890.
- [10] D. Hernikawati Balai Pengembangan Sumber Daya Manusia dan Penelitian Kominfo Jakarta, J. Pegangsaan Timur No, and J. Pusat, "Kecenderungan Tanggapan Masyarakat Terhadap Vaksin Sinovac Berdasarkan Lexicon Based Sentiment Analysis The Trend of Public Response to Sinovac Vaccine Based on Lexicon Based Sentiment Analysis," *Jurnal Ilmu Pengetahuan dan Teknologi Komunikasi*, vol. 23, no. 1, pp. 21–31, 2021, doi: 10.33169/iptekkom.23.1.2021.21-31.
- [11] F. Amaliah, I. Kadek, and D. Nuryana, "Perbandingan Akurasi Metode Lexicon Based Dan Naive Bayes Classifier Pada Analisis Sentimen Pendapat Masyarakat Terhadap Aplikasi Investasi Pada Media Twitter," *Journal of Informatics and Computer Science*, vol. 03, 2022.
- [12] F. Koto and G. Y. Rahmanyngtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," in *Proceedings of the 2017 International Conference on Asian Language Processing, IALP 2017*, Institute of Electrical and Electronics Engineers Inc., Jul. 2017, pp. 391–394. doi: 10.1109/IALP.2017.8300625.
- [13] G. E. Yulastuti, S. R. Wardhana, and A. Maulana, "Penerapan Insert-Lexicon untuk Analisis Sentimen Kenaikan Harga BBM pada Media Sosial Twitter," *Scan : Jurnal Teknologi Informasi dan Komunikasi*, vol. 19, no. 2, Jun. 2024, doi: 10.33005/scan.v19i2.4730.
- [14] R. Firdaus, I. A. #2, and A. Herdiani, "Lexicon-Based Sentiment Analysis of Indonesian Language Student Feedback Evaluation," 2019, doi: 10.34818/indojc.2021.6.1.408.
- [15] Y. Fauziah, B. Yuwono, and A. S. Aribowo, "Lexicon Based Sentiment Analysis in Indonesia Languages : A Systematic Literature Review," *malaysia*, 2021. [Online]. Available: <https://proceeding.researchsynergypress.com/index.php/cset/index>
- [16] Y. Nooryuda Prasetya and D. Winarso, "Penerapan Lexicon Based Untuk Analisis Sentimen Pada Twiter Terhadap Isu Covid-19," vol. 11, no. 2, 2021.
- [17] W. F. Abdillah, A. Premana, and R. M. Herdian Bhakti, "Analysis Of Covid-19 Handling Sentiment With Support Vector Machine: Lexicon Evaluation And Feature Extraction Methods," *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, vol. 3, no. 2, pp. 160–170, 2021.
- [18] M. K. Anam, T. A. Fitri, A. Agustin, L. Lusiana, M. B. Firdaus, and A. T. Nurhuda, "Sentiment Analysis for Online Learning using The Lexicon-Based Method and The Support Vector Machine Algorithm," *ILKOM Jurnal Ilmiah*, vol. 15, no. 2, pp. 290–302, Aug. 2023, doi: 10.33096/ilkom.v15i2.1590.290-302.
- [19] Sabika Amalina, "SENTIRESEARCH: LEXICON-BASED WEB APPLICATION FOR INDONESIAN SENTIMENT ANALYSIS," Fangshan, Nanjing, May 2020. Accessed: Sep. 03, 2025. [Online]. Available: <https://dspace.uui.ac.id/handle/123456789/23674;jsessionid=5BEC6F7B5F218AC7F095665572E41C35>