https://jurnal.pustakagalerimandiri.co.id/index.php/pustakaai
DOI : https://doi.org/10.55382/jurnalpustakaai.v5i2.1166

JURNAL PUSTAKA AI
JURNAL PUSAT AKSES KAJIAN TEKNOLOGI ARTIFICIAL INTELLIGENCE
Vol. 5 No. 2 (2025) 235 – 247
E ISSN: 2809-4069

Pemetaan Tren dan Pola Topik Kekerasan Remaja dengan ARM–TSA pada Pemberitaan *Online*

Nasyan Falah Azhari[1],Anita[*], Tresiana Pasaribu[3], Reinhard Halomoan Napitupulu [4]
[1234]Jurusan Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Prima Indonesia
[1]nasyanfalahazhari55@gmail.com, [*]anita@unprimdn.ac.id, [3]tresianapasaribu@gmail.com, [4]rennhrd.npl@gmail.com

Abstract

This study maps trends and topical patterns of adolescent violence in Indonesian online news (August 2024–May 2025) using a combination of Latent Dirichlet Allocation (LDA), Association Rule Mining (ARM), and Time Series Analysis (TSA). Data were collected via web scraping, yielding 5,272 articles. Preprocessing included tokenization, stop word removal, and stemming; LDA was optimized with a coherence score of 0.517 (K=10). Temporal analysis identified a reporting peak in April 2025 (847 articles). ARM found a strong association between "Sexual Violence" and "Child Protection" topics (lift > 1.2). The contributions are (1) an integrated LDA–ARM–TSA methodology for quantitatively mapping adolescent violence narratives and (2) temporal pattern findings relevant to institutional response. These results aim to support policymakers and child protection agencies in designing evidence-based interventions.

Keywords: Youth Violence, Online News Coverage, Web Scraping, Topic Modeling, Latent Dirichlet Allocation (LDA), Association Rule Mining (ARM), Time Series Analysis (TSA).

Abstrak

Penelitian ini memetakan tren dan pola topik kekerasan remaja dalam pemberitaan online di Indonesia (Agustus 2024–Mei 2025) menggunakan kombinasi *Latent Dirichlet Allocation* (LDA), Association Rule Mining (ARM), dan *Time Series Analysis* (TSA). Data diperoleh melalui web scraping sejumlah 5.272 artikel. Pra-pemrosesan meliputi tokenisasi, *stopword removal*, dan *stemming*; LDA dioptimalkan dengan coherence score = 0.517 (K=10). Analisis temporal menunjukkan puncak pemberitaan pada April 2025 (847 artikel). ARM mengungkap asosiasi kuat antara topik "Kekerasan Seksual" dan "Perlindungan Anak" (*lift* > 1.2). Kontribusi penelitian ini adalah integrasi metodologis LDA–ARM–TSA untuk memetakan narasi kekerasan remaja secara kuantitatif dan penemuan pola temporal yang berkaitan dengan respons institusional. Hasil ini diharapkan membantu pembuat kebijakan dan lembaga perlindungan anak dalam merancang intervensi berbasis bukti.

Kata kunci: Kekerasan Remaja, Pemberitaan *Online*, *Web Scraping*, Pemodelan Topik, *Latent Dirichlet Allocation* (LDA), *Association Rule Mining* (ARM), *Time Series Analysis* (TSA).

Submitted : 20-07-2025 | Reviewed : 25-08-2025 | Accepted : 02-08-2025

1. Pendahuluan

Kekerasan remaja merupakan fenomena sosial kompleks yang berdampak luas pada individu, keluarga, dan masyarakat. Fenomena ini mencakup kekerasan fisik, psikologis, seksual, hingga perilaku seperti *bullying* dan tawuran yang dapat terjadi di rumah, sekolah, maupun ruang publik. Faktor pemicunya meliputi pola asuh keluarga yang disfungsional, tekanan teman sebaya, paparan konten kekerasan di media, serta ketidakstabilan emosi [1]. Di era digital, media massa khususnya berita *online* yang memegang peran strategis dalam membentuk persepsi publik dan memengaruhi kebijakan sosial. Pemberitaan yang masif mengenai kekerasan remaja, seperti pada kasus OSPEK, kerap memicu *moral panic* di masyarakat serta memperkuat konstruksi sosial terhadap isu ini [2][3].

Sejumlah penelitian terdahulu telah membahas peran media dalam membingkai isu kekerasan remaja. Studi *framing* media menunjukkan bahwa media berperan dalam membentuk persepsi publik melalui pemilihan narasi, penggunaan bahasa, serta penekanan aspek tertentu dari suatu peristiwa [4][5]. Di sisi lain, pendekatan berbasis *Natural Language Processing* (NLP) telah digunakan untuk menganalisis teks berita secara otomatis, baik dalam mengekstraksi tema dominan maupun memetakan tren temporal [6][7]. Teknik *Topic Modeling*, khususnya algoritma *Latent Dirichlet Allocation* (LDA), terbukti efektif dalam mengidentifikasi topik dalam kumpulan berita berskala besar [8]. Sementara itu, *Association Rule Mining* (ARM) banyak digunakan untuk mengungkap pola keterkaitan antar kata kunci yang tersembunyi dalam data [9]. Namun, sebagian besar penelitian di Indonesia masih terbatas pada analisis manual atau mengandalkan data yang kurang luas dari segi jumlah dan rentang waktu, sehingga hasil yang diperoleh belum mampu memberikan gambaran yang komprehensif [10].

Penelitian ini dilakukan untuk menjawab keterbatasan tersebut dengan memanfaatkan pendekatan berbasis NLP dan data mining pada skala data yang lebih besar. Analisis manual tidak lagi memadai di tengah derasnya arus informasi dari media *online*. Gap penelitian terletak pada belum adanya kajian komprehensif yang memadukan analisis tren temporal, tema dominan, serta pola keterkaitan antar topik kekerasan remaja dengan pendekatan komputasional berbasis data. Kebaruan dari penelitian ini adalah penggunaan kombinasi metode *Time Series Analysis* (TSA), Topic Modeling, dan ARM untuk menganalisis pemberitaan kekerasan remaja dalam periode yang cukup panjang (Agustus 2024–Mei 2025), dengan fokus pada media berita *online* utama di Indonesia. Pendekatan ini diharapkan dapat memberikan perspektif baru dalam memahami bagaimana isu kekerasan remaja dikonstruksi dan diberitakan di ruang digital.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis tren volume pemberitaan kekerasan remaja di media online Indonesia selama periode Agustus 2024 hingga Mei 2025, mengidentifikasi topik-topik utama dalam pemberitaan kekerasan remaja menggunakan algoritma LDA, serta menemukan pola asosiasi antar kata kunci dan topik yang relevan melalui pendekatan ARM. Dengan demikian, hasil penelitian ini diharapkan dapat memberikan gambaran komprehensif mengenai narasi media terkait kekerasan remaja, sekaligus menjadi dasar untuk perumusan kebijakan pencegahan dan penanganan isu tersebut.

2. Metode Penelitian

2.1. Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis isi komputasional (computational content analysis). Dengan menggabungkan teknik-teknik dari bidang *Natural Language Processing (NLP)* dan *Data Mining*, penelitian ini bertujuan untuk mengidentifikasi pola-pola yang tidak dapat dengan mudah diamati melalui metode konvensional.

2.2. Populasi dan Sampel

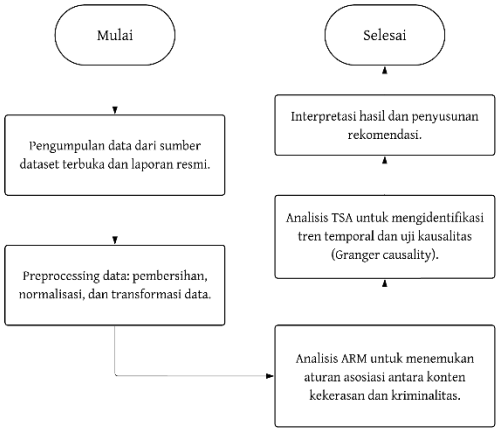
Populasi penelitian terdiri atas artikel berita online berbahasa Indonesia yang memuat topik "kekerasan remaja" dan diterbitkan oleh portal-portal berita nasional terkemuka. Sampel diambil menggunakan metode *non-probability sampling*, dengan pendekatan *convenience sampling* melalui teknik *web scraping*. Data artikel dikumpulkan dari beberapa portal berita seperti Kompas.com, Detik.com, CNNIndonesia.com, Liputan6.com, Tempo.co, dan Tribunnews.com. Rentang waktu pengambilan data adalah dari Agustus 2024 hingga Mei 2025, sehingga memungkinkan analisis tren temporal.

2.3. Instrumen Penelitian

Instrumen utama penelitian berupa kombinasi perangkat lunak dan pustaka Python untuk melakukan proses pengumpulan, pra-pemrosesan, dan analisis data. Analisis dijalankan di lingkungan *Google Colaboratory* (Colab), yang memungkinkan eksekusi kode berbasis *cloud*. Dalam pengambilan data digunakan *framework Scrapy* yang diadaptasi dari repositori binsarjr/news-scraper [11]. Beberapa library Python yang digunakan meliputi Pandas untuk manipulasi data, NLTK dan Sastrawi untuk tokenisasi, penghapusan stopwords, dan stemming,

BeautifulSoup4 untuk parsing HTML, mlxtend untuk algoritma Apriori, Gensim untuk LDA, serta Matplotlib dan Seaborn untuk visualisasi data. Untuk mempercepat eksekusi, pra-pemrosesan teks dijalankan dengan parallel processing menggunakan pandarallel.

2.4. Prosedur Penelitian



Gambar 1. Prosedur Penelitian

Prosedur penelitian terdiri dari lima tahap utama. Pertama, pengumpulan data dilakukan melalui *web scraping* pada portal berita dengan kata kunci "kekerasan remaja", dan hasilnya disimpan dalam format CSV [12]. Kedua, pra-pemrosesan teks meliputi pembersihan data dari duplikasi, *case folding*, penghapusan URL dan tag HTML, normalisasi *whitespace,* tokenisasi, *stopword removal*, dan *stemming*, dengan hasil pra-pemrosesan disimpan sebagai kolom baru dalam dataset. Ketiga, pemodelan topik dilakukan dengan metode *Latent Dirichlet Allocation* (LDA) yang dilatih pada corpus dengan format *Bag-of-Words* (BoW). Model ini dioptimalkan melalui *Coherence Score* untuk memastikan kualitas topik, dengan jumlah topik (K) yang ditetapkan sebesar 10 menggunakan rumus berikut:

$$y = f\,(x) = \frac{1}{(1+e)}\Delta F = -2{,}3\ x\ 10^{6}\ x\ F^{2}\frac{\Delta M}{A} \qquad P(w|d) = \sum_{k=1}^{K} P(w|z_k)P(z_k|d) \qquad (1)$$

Keempat, *Association Rule Mining* (ARM) digunakan untuk menemukan keterkaitan antar kata kunci atau antar topik dengan memanfaatkan algoritma Apriori [13]. Aturan asosiasi dievaluasi menggunakan tiga metrik kunci, yaitu *support*, *confidence*, dan *lift*. *Support (S)* mengukur frekuensi kemunculan itemset (kombinasi item) di seluruh transaksi, yang dituliskan sebagai:

$$S(X \cup Y) = \frac{\text{Jumlah transaksi mengandung X dan Y}}{\text{Jumlah total transaksi}} \qquad (2)$$

Confidence (C) mengukur probabilitas item YYY muncul ketika item XXX juga muncul, dengan rumus:

$$C(X \Rightarrow Y) = \frac{S(X \cup Y)}{S(X)} \qquad (3)$$

Sementara itu, *Lift (L)* menunjukkan sejauh mana kemunculan XXX meningkatkan probabilitas YYY muncul dibandingkan jika keduanya tidak terkait, dengan persamaan:

$$L(X \Rightarrow Y) = \frac{S(X \cup Y)}{S(X) \times S(Y)} \qquad (4)$$

Nilai L > 1 menunjukkan adanya asosiasi positif antara item X dan Y. Kelima, *Time Series Analysis* (TSA) diterapkan untuk memetakan tren jumlah pemberitaan dan tren per topik secara temporal, baik dalam skala harian, mingguan, maupun bulanan, dengan visualisasi berupa grafik garis [16][17].

2.5. Teknik Analisis Data

Teknik analisis data yang digunakan meliputi analisis deskriptif untuk memahami distribusi data, analisis kuantitatif berbasis teks, analisis topik dengan LDA, ARM untuk menemukan pola keterkaitan, serta analisis deret waktu untuk melihat dinamika tren. Temuan-temuan kuantitatif ini kemudian disintesis secara kualitatif untuk memberikan pemahaman mendalam tentang konteks pemberitaan kekerasan remaja. Pendekatan komputasional

yang dijelaskan di atas dirancang agar dapat direplikasi dengan parameter, *pipeline*, dan pustaka yang sama, sesuai dengan standar keterulangan penelitian modern [14].

3. Hasil dan Pembahasan

3.1 Deskripsi Data Hasil Pengumpulan

```
query_keywords = "kekerasan remaja"
start_date = "2024-08-01"
end_date = "2025-05-31"
output_file = "/content/drive/My Drive/data_artikel_kekerasanremaja_scrappy.csv"

from google.colab import drive
drive.mount('/content/drive')

!poetry run python main.py -q "{query_keywords}" --since "{start_date}" --until "{end_date}" --output "{output_file}"

print("Proses scraping selesai. Data Anda seharusnya sudah tersimpan di Google Drive.")
```

Gambar 2. Kode Python Proses Scrapping Data

```
Contoh 5 baris pertama data yang berhasil di-scrape:
  author                                            content           keyword  \
0    NaN  judul: \n\n\n\n\n\nauthor: \ntanggal: Selasa, ...  kekerasan remaja
1    NaN  judul: \n\n\n\n\n\nauthor: \ntanggal: Rabu, 28...  kekerasan remaja
2    NaN  judul: \n\n\n\n\n\nauthor: \ntanggal: Selasa, ...  kekerasan remaja
3    NaN  judul: \n\n\n\n\n\nauthor: \ntanggal: Sabtu, 2...  kekerasan remaja
4    NaN  judul: \n\n\n\n\n\nauthor: \ntanggal: Minggu, ...  kekerasan remaja

                                                link         publish_date  \
0  https://www.cnnindonesia.com/nasional/20250526...  2025-05-27 06:33:00
1  https://www.cnnindonesia.com/hiburan/202505261...  2025-05-28 09:30:00
2  https://www.cnnindonesia.com/internasional/202...  2025-05-27 07:42:00
3  https://www.cnnindonesia.com/gaya-hidup/202505...  2025-05-24 13:30:00
4  https://www.cnnindonesia.com/nasional/20250518...  2025-05-18 15:20:00

          source       title
0  cnnindonesia.com  \n\n\n\n\n
1  cnnindonesia.com  \n\n\n\n\n
2  cnnindonesia.com  \n\n\n\n\n
3  cnnindonesia.com  \n\n\n\n\n
4  cnnindonesia.com  \n\n\n\n\n

Total artikel di file CSV: 5274
```

Gambar 3. Contoh Data yang Berhasil di-Scrape

Setelah proses *crawling* dan pengumpulan, diperoleh total 5272 dokumen (artikel) yang relevan yang masih berupa data mentah.

3.2 Hasil Pra-pemrosesan Data

```
if 'keyword' in df.columns:
    df = df.drop(columns=['keyword'])
    print("\nKolom 'keyword' dihapus.")
df['publish_date'] = pd.to_datetime(df['publish_date'], errors='coerce')
print("Kolom 'publish_date' dikonversi ke tipe datetime.")

initial_rows = len(df)
df.dropna(subset=['link', 'content', 'title', 'publish_date'], inplace=True)
print(f"Baris dengan link, content, title, atau publish_date kosong dihapus. Sisa: {len(df)} baris.")

df.drop_duplicates(subset=['link'], inplace=True)
print(f"Duplikasi berdasarkan link dihapus. Sisa: {len(df)} baris.")
print(f"Total baris setelah pembersihan awal: {len(df)}")
```

Gambar 4. Kode Python Pembersihan Awal Data

```
preprocess_text_and_measure_time(text):
start_time = time.time() # Mulai mengukur waktu

if pd.isna(text) or not isinstance(text, str):
    return "", 0.0 # Kembalikan string kosong dan waktu 0.0

processed_text = text.lower()

processed_text = re.sub(r'http\S+|www\S+|https\S+', '', processed_text, flags=re.MULTILINE)

processed_text = BeautifulSoup(processed_text, 'html.parser').get_text()

processed_text = re.sub(r'[^a-zA-Z\s]', '', processed_text)
processed_text = re.sub(r'\d+', '', processed_text)
processed_text = re.sub(r'\s+', ' ', processed_text).strip()

tokens = word_tokenize(processed_text)
tokens = [word for word in tokens if word not in indonesian_stopwords]
tokens = [stemmer.stem(word) for word in tokens]

end_time = time.time() # Selesai mengukur waktu
return ' '.join(tokens), (end_time - start_time) # Kembalikan teks dan waktu
```

Gambar 5. Kode Python Pra-pemrosesan lanjutan

Pembersihan Awal Data, *Case Folding*, Penghapusan URL dan Tag HTML, Penghapusan Karakter non-Alfabet dan Angka, Normalisasi *Whitespace*, Tokenisasi, *Stopword Removal*, *Stemming*. Setelah semua tahapan ini, total 5272 artikel tetap dipertahankan, menunjukkan kualitas data yang baik setelah pembersihan awal.

```
Contoh 5 baris pertama data setelah pra-pemrosesan (dengan waktu):
       title processed_title  \
0  \n\n\n\n\n
1  \n\n\n\n\n
2  \n\n\n\n\n
3  \n\n\n\n\n
4  \n\n\n\n\n

                                             content  \
0  judul: \n\n\n\n\n\nauthor: \ntanggal: Selasa, ...
1  judul: \n\n\n\n\n\nauthor: \ntanggal: Rabu, 28...
2  judul: \n\n\n\n\n\nauthor: \ntanggal: Selasa, ...
3  judul: \n\n\n\n\n\nauthor: \ntanggal: Sabtu, 2...
4  judul: \n\n\n\n\n\nauthor: \ntanggal: Minggu, ...

                                   processed_content  \
0  judul author tanggal selasa mei wib jakarta cn...
1  judul author tanggal rabu mei wib jakarta cnn ...
2  judul author tanggal selasa mei wib jakarta cn...
3  judul author tanggal sabtu mei wib daftar isi ...
4  judul author tanggal minggu mei wib jakarta cn...
```

Gambar 6. Contoh Data Hasil Pra-pemrosesan

3.3 Hasil Pemodelan Topik (*Topic Modeling*)

```python
texts = [doc.split() for doc in df_processed['combined_processed_text'] if isinstance(doc, str)]

dictionary = corpora.Dictionary(texts)

dictionary.filter_extremes(no_below=5, no_above=0.5, keep_n=100000)

corpus = [dictionary.doc2bow(text) for text in texts]
```

Gambar 7. Kode Python Fungsi Corpus dan Vocabulary

```python
lda_model = LdaModel(
    corpus=corpus,
    id2word=dictionary,
    num_topics=num_topics,
    random_state=100,
    update_every=1,
    chunksize=100,
    passes=10,
    alpha='auto',
    eta='auto',
    per_word_topics=True
)
```

Gambar 8. Kode Python Pemodelan LDA

Analisis pemodelan topik bertujuan untuk mengidentifikasi tema-tema yang mendasari kumpulan artikel, memberikan gambaran abstrak tentang fokus pemberitaan.

a. *Corpus* dan *Vocabulary*

Model LDA dilatih menggunakan corpus yang terdiri dari 5272 dokumen (artikel). etelah proses pemfilteran, vocabulary yang digunakan dalam model berjumlah 5000 kata unik.

```
Jumlah dokumen (artikel) dalam corpus: 5272
Jumlah kata unik (vocabulary) setelah filter: 10244

Contoh representasi Bag-of-Words untuk dokumen pertama:
[(0, 2), (1, 5), (2, 5), (3, 1), (4, 1), (5, 1), (6, 5), (7, 1), (8, 1), (9, 2), (10, 1), (11, 37), (12, 2), (13, 1), (14, 2), (15, 3), (16, 11),

Beberapa mapping kata ke ID dari dictionary:
[(0, 'acu'), (1, 'adil'), (2, 'agama'), (3, 'agung'), (4, 'agustinus'), (5, 'ahli'), (6, 'aju'), (7, 'akhir'), (8, 'aksesibilitas'), (9, 'alami')]
```

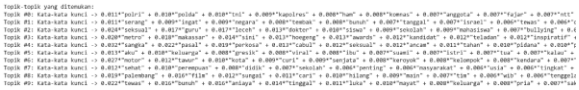
Gambar 9. Hasil Output Corpus dan Vocabulary

b. Pemodelan LDA

Model *Latent Dirichlet Allocation* (LDA) dilatih dengan menentukan jumlah 10 topik. LDA mengasumsikan bahwa setiap dokumen adalah campuran dari berbagai topik, dan setiap topik adalah campuran dari berbagai kata. Distribusi ini dinyatakan dengan P(w|zk) sebagai distribusi kata dalam topik zkz, dan P(zk|d) sebagai distribusi topik dalam dokumen. Perhitungan probabilitas gabungan kata dan topik ini didasarkan pada rumus persamaan (1) yang digunakan dalam model LDA untuk mengestimasi parameter distribusi topik dan kata.

c. Interpretasi Topik

Setiap topik diidentifikasi melalui daftar kata-kata kunci teratas yang memiliki probabilitas tertinggi untuk muncul dalam topik tersebut.



Gambar 10. Topik-Topik yang ditemukan untuk Interpretasi

Berdasarkan kata-kata kunci ini, interpretasi dan penamaan topik dilakukan sebagai berikut,

```python
def get_document_topics(lda_model, corpus):
    # Mengambil topik dan probabilitasnya untuk setiap dokumen
    document_topics = []
    for i, row in enumerate(lda_model[corpus]):
        # Urutkan topik berdasarkan probabilitas (descending)
        row = sorted(row[0], key=lambda x: x[1], reverse=True)

        # Ambil topik dominan
        topic_num = row[0][0]
        prop_topic = row[0][1]
        wp = lda_model.show_topic(topic_num, topn=5)
        topic_keywords = ", ".join([word for word, prop in wp])

        document_topics.append({
            'Document_ID': i,
            'Dominant_Topic': topic_num,
            'Perc_Contribution': round(prop_topic, 4),
            'Topic_Keywords': topic_keywords
        })
    return pd.DataFrame(document_topics)

df_topic_distribution = get_document_topics(lda_model, corpus)
```

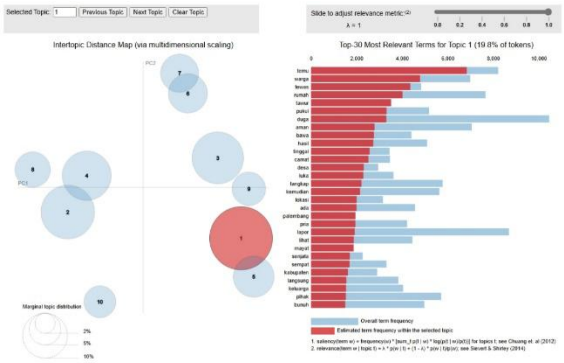
Gambar 11. Kode Python Fungsi Interpretasi dan Penamaan Topik

```
Contoh 5 baris pertama DataFrame dengan topik dominan:
       title  Dominant_Topic  Perc_Contribution  \
0  \n\n\n\n\n               3             0.7118
1  \n\n\n\n\n               9             0.4451
2  \n\n\n\n\n               2             0.5135
3  \n\n\n\n\n               3             0.8582
4  \n\n\n\n\n               5             0.3743

                        Topic_Keywords
0       anak, sehat, lebih, usia, sosial
1      didik, sekolah, guru, siswa, ajar
2  masyarakat, tindak, hukum, aman, kota
3       anak, sehat, lebih, usia, sosial
4        lapor, anak, duga, jakarta, ibu
```

Gambar 12. Output Contoh 5 Baris Pertama DataFrame dengan Topik Dominan

d. Visualisasi Kualitas Model

Visualisasi menggunakan pyLDAvis memberikan representasi interaktif, serta membantu dalam mengevaluasi pemisahan antar topik dan relevansi kata kunci.



Gambar 13. Visualisasi Kualitas Model

e. Coherence Score

```
Coherence Score Model LDA (c_v): 0.45397913879529395
```

Gambar 14. Coherence Score

Coherence Score model LDA yang diperoleh adalah 0.5172359708156267. Nilai berikut sudah berada diatas rata-rata nilai coherence model LDA, yaitu 0.38. Hal ini menunjukkan bahwa hasil model dalam penelitian ini telah mencapai kualitas interpretatif yang layak dan dapat diterima dalam praktik [15].

3.4 Hasil *Association Rule Mining* (ARM)

Analisis ini bertujuan untuk mengidentifikasi asosiasi granular antara kata-kata yang telah diproses dalam artikel.

```
min_confidence_word = 0.5
min_lift_word = 1.2

print(f"Menghasilkan aturan asosiasi kata dengan min_confidence = {min_confidence_word} dan min_lift = {min_lift_word}...")
rules_word = association_rules(frequent_itemsets_word, metric="lift", min_threshold=min_lift_word)

rules_word = rules_word[rules_word['confidence'] >= min_confidence_word]
rules_word = rules_word.sort_values(by='lift', ascending=False)

print(f"Ditemukan {len(rules_word)} aturan asosiasi kata.")
print("\nContoh 10 aturan asosiasi kata teratas (berdasarkan lift):")
print(rules_word.head(10))
```

Gambar 15. Kode Python fungsi Data Transaksi

```
if 'df_transactions_word' in locals() and not df_transactions_word.empty:
    min_support_word = 0.3

    print(f"Mencari frequent itemsets kata dengan min_support = {min_support_word}...")
    frequent_itemsets_word = apriori(df_transactions_word, min_support=min_support_word, use_colnames=True)

    frequent_itemsets_word['length'] = frequent_itemsets_word['itemsets'].apply(lambda x: len(x))
    frequent_itemsets_word = frequent_itemsets_word.sort_values(by='support', ascending=False)
```

Gambar 16. Kode Python fungsi *Frequent Itemset*

```
if 'frequent_itemsets_word' in locals() and not frequent_itemsets_word.empty:
    min_confidence_word = 0.5
    min_lift_word = 1.2

    print(f"Menghasilkan aturan asosiasi kata dengan min_confidence = {min_confidence_word} dan min_lift = {min_lift_word}...")
    rules_word = association_rules(frequent_itemsets_word, metric="lift", min_threshold=min_lift_word)

    rules_word = rules_word[rules_word['confidence'] >= min_confidence_word]
    rules_word = rules_word.sort_values(by='lift', ascending=False)
```

Gambar 17. Kode Python fungsi Aturan Asosiasi Kata

Data teks yang didapat dari hasil proses sebelumnya diubah menjadi format transaksi one-hot encoded. Jumlah transaksi (artikel) adalah 5272, dan jumlah kata unik (item) yang digunakan setelah filter agresif adalah 79196 kata unik.

```
Jumlah transaksi (artikel) untuk ARM berbasis kata: 5272
Jumlah item unik (kata): 79196

Contoh 5 baris pertama data dalam format transaksi kata (One-Hot Encoded):
       a     aa  aaaryan    aaf    aag  aagkronologi    aah  aahbaca  \
0  False  False    False  False  False         False  False    False
1  False  False    False  False  False         False  False    False
2  False  False    False  False  False         False  False    False
3  False  False    False  False  False         False  False    False
4   True  False    False  False  False         False  False    False

   aahketiga  aahpenelusuran  ...  zulkarnainbaca  zulkarnainkecurigaan  \
0      False           False  ...           False                 False
1      False           False  ...           False                 False
2      False           False  ...           False                 False
3      False           False  ...           False                 False
4      False           False  ...           False                 False

   zulkarnainzulkarnain  zulkifli  zulkifri  zurich  zurriyatun  zwart  \
0                 False     False     False   False       False  False
1                 False     False     False   False       False  False
2                 False     False     False   False       False  False
3                 False     False     False   False       False  False
4                 False     False     False   False       False  False
```

Gambar 18. Contoh Transaksi Artikel yang didapat

Algoritma Apriori digunakan untuk menemukan *frequent itemsets* dengan min_support = 0.3. Total 44210 dokumen digunakan sebagai transaksi. Didapatkan sejumlah besar *frequent itemsets*.

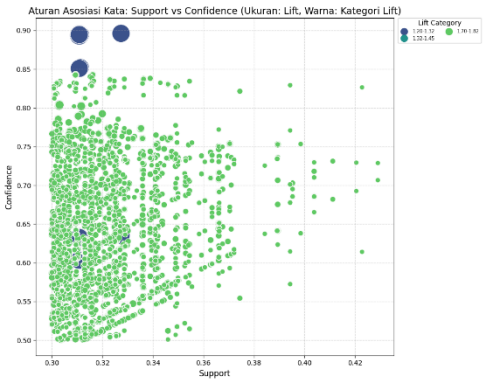
```
Ditemukan 44210 frequent itemsets kata.

Contoh 10 frequent itemsets kata teratas:
       support                              itemsets  length
73     0.949734                               (with)       1
59     0.949355                             (scroll)       1
14542  0.949165  (to, with, content, continue, scroll)       5
267    0.949165                      (content, to)       2
694    0.949165                         (to, with)       2
6767   0.949165           (content, scroll, to, with)       4
258    0.949165                    (content, scroll)       2
2168   0.949165                (continue, scroll, to)       3
311    0.949165                    (continue, scroll)       2
1856   0.949165                 (content, to, with)       3
```

Gambar 19. Contoh *Frequent Itemset* Data yang dihasilkan

Aturan asosiasi di-*generate* dari *frequent itemsets* dengan metrik *lift* minimal 1.2 dan *confidence* minimal 0.5. Jumlah aturan asosiasi yang ditemukan adalah 201474.

```
Contoh 10 Aturan Asosiasi Kata Teratas (berdasarkan Lift):
                                antecedents  \
0     frozenset({'satu', 'content', 'with'})
1      frozenset({'satu', 'scroll', 'with'})
2           frozenset({'continue', 'salah'})
3   frozenset({'content', 'scroll', 'salah'})
4    frozenset({'satu', 'content', 'scroll'})
5              frozenset({'satu', 'scroll'})
6             frozenset({'scroll', 'salah'})
7       frozenset({'scroll', 'to', 'salah'})
8        frozenset({'satu', 'scroll', 'to'})
9      frozenset({'content', 'to', 'salah'})
```

Gambar 20. Contoh Aturan Asosiasi Kata yang dihasilkan



Gambar 21. *Scatter Plot* Aturan Asosiasi Kata: *Support vs Confidence*

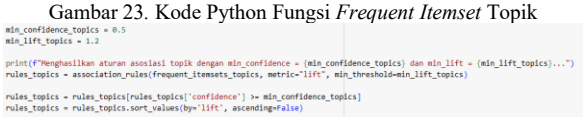
Analisis ARM berbasis topik memberikan gambaran tingkat tinggi tentang bagaimana tema-tema kekerasan remaja saling berhubungan.

```python
if 'lda_model' in locals() and 'corpus' in locals() and not df_processed.empty:
    topic_contribution_threshold = 0.05

    list_of_topic_transactions = []
    for i, doc_bow in enumerate(corpus):
        document_topics = lda_model.get_document_topics(doc_bow, minimum_probability=None)

        relevant_topics = []
        for topic_id, prop in document_topics:
            if prop >= topic_contribution_threshold:
                relevant_topics.append(f"Topic_{topic_id}")

        if relevant_topics:
            list_of_topic_transactions.append(relevant_topics)

    te_topic = TransactionEncoder()
    te_ary_topics = te_topic.fit(list_of_topic_transactions).transform(list_of_topic_transactions)
    df_transactions_topics = pd.DataFrame(te_ary_topics, columns=te_topic.columns_)
```

Gambar 22. Kode Python Fungsi Data Transaksi Topik

```
if 'df_transactions_topics' in locals() and not df_transactions_topics.empty:
    min_support_topics = 0.01

    print(f"Mencari frequent itemsets topik dengan min_support = {min_support_topics}...")
    frequent_itemsets_topics = apriori(df_transactions_topics, min_support=min_support_topics, use_colnames=True)

    frequent_itemsets_topics['length'] = frequent_itemsets_topics['itemsets'].apply(lambda x: len(x))
    frequent_itemsets_topics = frequent_itemsets_topics.sort_values(by='support', ascending=False)

    print(f"Ditemukan {len(frequent_itemsets_topics)} frequent itemsets topik.")
    print("\nContoh 10 frequent itemsets topik teratas:")
    print(frequent_itemsets_topics.head(10))

    output_frequent_itemsets_topics_csv_path = '/content/drive/My Drive/frequent_itemsets_topics_fase3.csv'
```

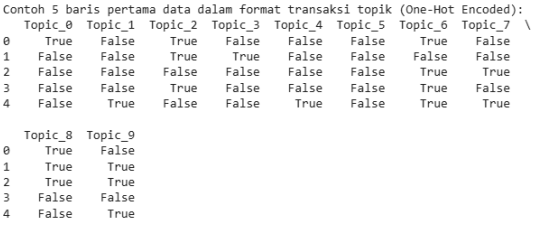
Gambar 23. Kode Python Fungsi *Frequent Itemset* Topik



```
min_confidence_topics = 0.5
min_lift_topics = 1.2

print(f"Menghasilkan aturan asosiasi topik dengan min_confidence = {min_confidence_topics} dan min_lift = {min_lift_topics}...")
rules_topics = association_rules(frequent_itemsets_topics, metric="lift", min_threshold=min_lift_topics)

rules_topics = rules_topics[rules_topics['confidence'] >= min_confidence_topics]
rules_topics = rules_topics.sort_values(by='lift', ascending=False)
```

Gambar 24. Kode Python Fungsi Aturan Asosiasi Topik

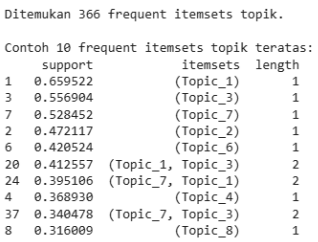
Setiap artikel direpresentasikan sebagai transaksi yang berisi daftar topik (Topic_ID) yang memiliki kontribusi signifikan (ambang batas kontribusi topik 0.05) pada artikel tersebut.



```
Contoh 5 baris pertama data dalam format transaksi topik (One-Hot Encoded):
   Topic_0  Topic_1  Topic_2  Topic_3  Topic_4  Topic_5  Topic_6  Topic_7  \
0     True    False     True    False    False    False     True    False
1    False    False     True     True    False    False    False    False
2    False    False    False    False    False    False     True     True
3    False    False     True    False    False    False     True    False
4    False     True    False    False     True    False     True     True

   Topic_8  Topic_9
0     True    False
1     True     True
2     True     True
3    False    False
4    False     True
```

Gambar 25. Contoh Data dalam Format Transaksi Topik (*One-Hot Encoded*)

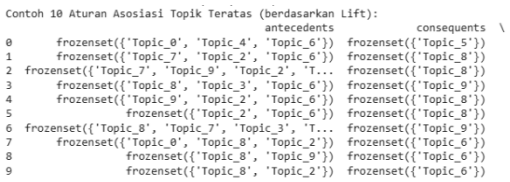
Algoritma Apriori diterapkan pada transaksi topik dengan min_support = 0.01. Jumlah *frequent itemsets* topik yang ditemukan adalah sebanyak 366 *frequent itemsets* topik.



```
Ditemukan 366 frequent itemsets topik.

Contoh 10 frequent itemsets topik teratas:
     support            itemsets  length
1   0.659522           (Topic_1)       1
3   0.556904           (Topic_3)       1
7   0.528452           (Topic_7)       1
2   0.472117           (Topic_2)       1
6   0.420524           (Topic_6)       1
20  0.412557  (Topic_1, Topic_3)       2
24  0.395106  (Topic_7, Topic_1)       2
4   0.368930           (Topic_4)       1
37  0.340478  (Topic_7, Topic_3)       2
8   0.316009           (Topic_8)       1
```

Gambar 26. Contoh *Frequent Itemsets* Topik yang dihasilkan

Aturan asosiasi topik di-*generate* dari *frequent itemset*s topik dengan min_confidence = 0.5 dan min_lift = 1.2.



```
Contoh 10 Aturan Asosiasi Topik Teratas (berdasarkan Lift):
                                         antecedents                consequents  \
0       frozenset({'Topic_0', 'Topic_4', 'Topic_6'})  frozenset({'Topic_5'})
1       frozenset({'Topic_7', 'Topic_2', 'Topic_6'})  frozenset({'Topic_8'})
2  frozenset({'Topic_7', 'Topic_9', 'Topic_2', 'T...  frozenset({'Topic_8'})
3       frozenset({'Topic_8', 'Topic_3', 'Topic_6'})  frozenset({'Topic_9'})
4       frozenset({'Topic_9', 'Topic_2', 'Topic_6'})  frozenset({'Topic_8'})
5                 frozenset({'Topic_2', 'Topic_6'})  frozenset({'Topic_8'})
6  frozenset({'Topic_8', 'Topic_7', 'Topic_3', 'T...  frozenset({'Topic_9'})
7       frozenset({'Topic_0', 'Topic_8', 'Topic_2'})  frozenset({'Topic_6'})
8                 frozenset({'Topic_8', 'Topic_9'})  frozenset({'Topic_6'})
9                 frozenset({'Topic_8', 'Topic_2'})  frozenset({'Topic_6'})
```

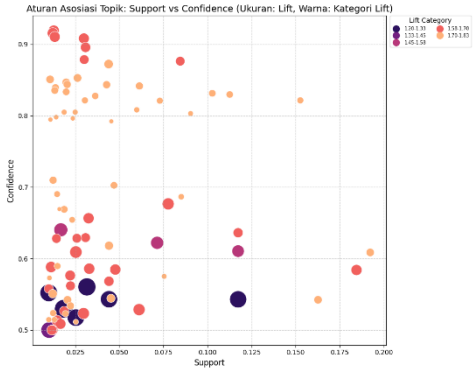
Gambar 27. Contoh Aturan Asosiasi Topik teratas yang dihasilkan (*antecendents, consequents*)

	consequent support	support	confidence	lift	representativity	\
328	0.290212	0.018589	0.529730	1.825317	1.0	
164	0.316009	0.031677	0.560403	1.773375	1.0	
645	0.316009	0.010053	0.552083	1.747049	1.0	
225	0.300076	0.025417	0.517375	1.724146	1.0	
98	0.316009	0.044196	0.543124	1.718696	1.0	
11	0.316009	0.117413	0.542982	1.718249	1.0	
610	0.300076	0.010243	0.500000	1.666245	1.0	
354	0.420524	0.016882	0.640288	1.522597	1.0	
44	0.420524	0.071510	0.622112	1.479376	1.0	
12	0.420524	0.117413	0.610454	1.451652	1.0	

	leverage	conviction	zhangs_metric	jaccard	certainty	kulczynski
328	0.008405	1.509318	0.468593	0.060606	0.337449	0.296891
164	0.013814	1.555949	0.462231	0.092933	0.357305	0.330321
645	0.004299	1.527049	0.435537	0.031012	0.345142	0.291948
225	0.010675	1.450243	0.441702	0.078500	0.310460	0.301039
98	0.018481	1.497103	0.455205	0.125134	0.332043	0.341490
11	0.049080	1.496640	0.533340	0.283036	0.331837	0.457266
610	0.004096	1.399848	0.408211	0.033007	0.285637	0.267067
354	0.005794	1.610945	0.352522	0.039259	0.379246	0.340216
44	0.023172	1.533462	0.366124	0.154129	0.347881	0.396081
12	0.036531	1.487567	0.385222	0.236983	0.327762	0.444830

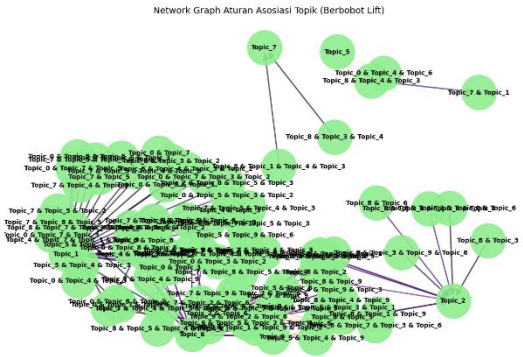
Gambar 28. Contoh Aturan Asosiasi Topik teratas yang dihasilkan (*Consequent support, support, confidence, lift representativity, leverage, conviction, zhang_metric, jaccard, certainity,kulcynski*)

Jumlah aturan asosiasi topik yang ditemukan adalah sebanyak 85 aturan asosiasi topik.



Gambar 29. *Scatter Plot* Aturan Asosiasi Topik*: Support vs Confidence*

Gambar di atas menampilkan distribusi aturan asosiasi topik yang diperoleh dari analisis berbasis topik. Setiap titik pada *scatter plot* merepresentasikan satu aturan asosiasi antara topik-topik tertentu dengan menggunakan tiga metrik utama, yaitu *support, confidence*, dan *lift*. Sumbu X menunjukkan nilai *support,* yang menggambarkan seberapa sering kombinasi topik tersebut muncul secara bersamaan. Sumbu Y menunjukkan nilai *confidence*, yang merepresentasikan kekuatan hubungan antar topik. Ukuran titik pada plot mencerminkan nilai *lift*, di mana semakin besar ukuran titik, semakin kuat asosiasi yang ditunjukkan, dengan nilai lebih dari 1,0 mengindikasikan asosiasi yang signifikan. Selain itu, warna titik digunakan untuk membedakan kategori *lift*, sehingga mempermudah identifikasi tingkat kekuatan asosiasi secara visual. Visualisasi ini membantu peneliti dalam mengenali aturan yang paling relevan dan memiliki keterkaitan yang kuat dalam konteks pemberitaan kekerasan remaja.

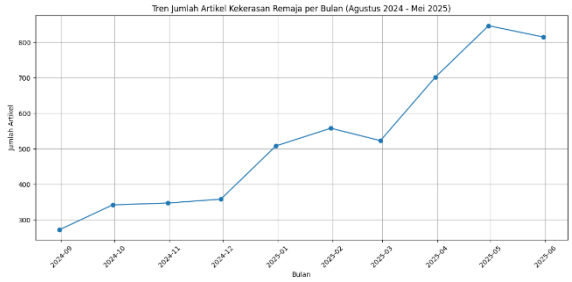


Gambar 30. *Network Graph* Aturan Asosiasi Topik

Berdasarkan gambar diatas, grafik ini memvisualisasikan hubungan antar topik yang ditemukan dari analisis Association Rule Mining (ARM) berbasis topik. Ini adalah representasi visual yang sangat efektif untuk memahami pola keterkaitan antara tema-tema utama dalam pemberitaan kekerasan remaja.

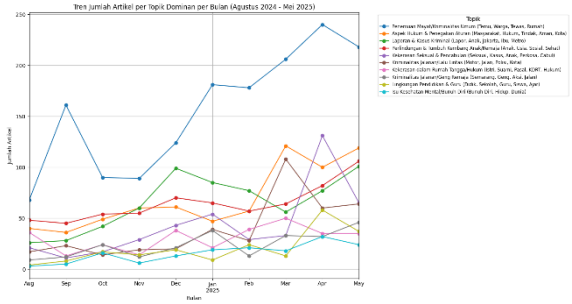
3.5 Sintesis Temuan (Integrasi ARM dan TSA)

Integrasi lintas pendekatan ini memberikan pemahaman yang lebih komprehensif mengenai pemberitaan, tidak hanya terkait dengan apa yang diberitakan, tetapi juga kapan isu-isu tertentu mendapatkan sorotan besar serta bagaimana topik-topik tersebut saling dikaitkan secara naratif dalam media. Analisis tren menunjukkan bahwa puncak pemberitaan terjadi pada April 2025 dengan total 847 artikel, yang sebagian besar berfokus pada topik Kriminalitas Jalanan, Kekerasan dalam Rumah Tangga, dan Kekerasan Seksual. Tren ini berlanjut pada Mei 2025 dengan 815 artikel, mencerminkan konsistensi perhatian media terhadap isu-isu publik yang bernuansa kekerasan dan berkaitan erat dengan penegakan hukum.



Gambar 31. Tren Jumlah Artikel Kekerasan Remaja per Bulan

Pada Maret 2025, tercatat sebanyak 702 artikel dengan dominasi topik yang serupa, menunjukkan keberlanjutan kejadian dan intensitas respons media selama tiga bulan berturut-turut. Pola ini mengindikasikan tingginya sensitivitas media online terhadap insiden aktual dan viral, serta kecenderungan *framing* narasi kekerasan yang menyoroti peran pelaku, korban, dan institusi hukum. Analisis *Association Rule Mining* (ARM) mengungkap pola naratif dominan dalam pemberitaan, di mana hasil ARM berbasis kata menunjukkan konsistensi tema melalui kemunculan kata kunci utama seperti “polisi”, “tangkap”, “pelaku”, dan “lapor” yang secara berulang menjadi fokus dalam liputan kekerasan remaja.



Gambar 32. Tren Jumlah Artikel per Topik Dominan

Secara temporal, terdapat pergeseran narasi: awal periode (Agustus–Oktober 2024) didominasi isu *bullying* dan kekerasan sekolah, sementara sejak Januari 2025 fokus bergeser ke respon hukum dan institusional, mencerminkan transisi dari narasi preventif ke represif.

Temuan dari integrasi pemodelan topik, analisis tren waktu, dan asosiasi antar topik maupun kata menunjukkan bahwa pemberitaan kekerasan remaja di media online Indonesia tidak berlangsung secara acak, melainkan membentuk pola tematik dan naratif yang terstruktur. Media online cenderung menyoroti insiden kekerasan remaja secara intensif ketika terjadi peristiwa besar atau viral, kemudian memperluas narasi dengan mengaitkannya pada aspek hukum, pendidikan, perlindungan anak, dan kesehatan mental. Hal ini memperlihatkan bahwa media berperan ganda, yakni sebagai pelapor fakta sekaligus pembentuk persepsi publik mengenai kekerasan remaja sebagai isu sosial yang kompleks. Secara keseluruhan, narasi media selama periode Agustus 2024 hingga Mei 2025 menunjukkan pergeseran fokus dari kekerasan ringan di awal periode, seperti bullying atau tawuran, menuju isu kekerasan yang lebih serius dan berskala hukum, seperti kekerasan seksual dan kriminalitas jalanan, pada paruh

akhir periode. Pergeseran ini dapat mencerminkan dinamika sosial yang berkembang maupun strategi editorial media dalam menyesuaikan pemberitaan dengan atensi publik dan kebijakan otoritas.

4. Kesimpulan

Penelitian ini mengungkap bahwa pemberitaan mengenai kekerasan remaja di media online Indonesia mengalami fluktuasi signifikan dengan puncak pada periode tertentu ketika isu tersebut menjadi viral atau mendapat perhatian nasional. Analisis topik menemukan sepuluh tema utama, dengan dominasi pada isu kriminalitas jalanan, geng remaja, kekerasan seksual, pencabulan, serta aspek hukum. Terdapat pergeseran fokus narasi dari kekerasan ringan, seperti bullying dan tawuran, menuju isu kekerasan yang lebih serius dan berdampak hukum pada paruh akhir periode analisis. Melalui pendekatan kombinasi LDA, ARM, dan TSA, ditemukan adanya keterkaitan tematik yang kuat antara kekerasan seksual dengan perlindungan anak, serta kekerasan jalanan dengan penegakan hukum. Selain itu, kata-kata seperti "polisi", "tangkap", "korban", dan "lapor" menjadi penanda utama narasi yang berulang dalam pemberitaan. Temuan ini menunjukkan bahwa integrasi analisis berbasis NLP dan data mining efektif untuk mengungkap pola isi, hubungan antar topik, serta dinamika temporal pemberitaan kekerasan remaja. Aplikasi hasil penelitian ini dapat mendukung media massa dalam merancang pemberitaan yang lebih berimbang, membantu pembuat kebijakan dalam merumuskan strategi pencegahan kekerasan remaja, serta menjadi dasar penguatan peran lembaga perlindungan anak dalam merespons isu sosial tersebut.

Penelitian selanjutnya disarankan untuk menggabungkan analisis sentimen guna menggali sikap media terhadap isu kekerasan remaja secara lebih mendalam. Penggunaan sumber data tambahan, seperti media sosial atau forum daring, juga dapat memberikan perspektif perbandingan antara narasi publik dan media arus utama. Kajian lanjutan sebaiknya meneliti hubungan antara lonjakan pemberitaan dengan kebijakan atau peristiwa sosial tertentu, serta mengelompokkan data berdasarkan jenis kekerasan atau periode waktu untuk memperkuat konteks analisis. Hasil penelitian ini dapat dimanfaatkan oleh pemerintah, lembaga perlindungan anak, dan media massa sebagai dasar untuk menyusun strategi pencegahan, intervensi, serta praktik jurnalisme yang lebih solutif dan berimbang.

Daftar Rujukan

[1] D. Ningrum, R. M. A. Hutagaol, and A. Muhtar, "Penerapan Time Series Forecasting untuk Memprediksi Pertumbuhan Ekonomi Indonesia 2024," *Data Sciences Indonesia (DSI)*, vol. 3, no. 2, pp. 79–89, May 2024, doi: 10.47709/dsi.v3i2.3263.

[2] L. Michaud, G. Kolla, K. Rudzinski, and A. Guta, "Mapping a moral panic: News media narratives and medical expertise in public debates on safer supply, diversion, and youth drug use in Canada," *International Journal of Drug Policy*, vol. 127, p. 104423, 2024, doi: https://doi.org/10.1016/j.drugpo.2024.104423.

[3] Supriyono, A. P. Wibawa, Suyono, and F. Kurniawan, "Advancements in natural language processing: Implications, challenges, and future directions," *Telematics and Informatics Reports*, vol. 16, p. 100173, 2024, doi: https://doi.org/10.1016/j.teler.2024.100173.

[4] S. R. Muir, L. D. Roberts, and L. P. Sheridan, "The portrayal of online shaming in contemporary online news media: A media framing analysis," *Computers in Human Behavior Reports*, vol. 3, p. 100051, 2021, doi: https://doi.org/10.1016/j.chbr.2020.100051.

[5] M. H. Santoso, "Application of Association Rule Method Using Apriori Algorithm to Find Sales Patterns Case Study of Indomaret Tanjung Anom," *Brilliance: Research of Artificial Intelligence*, vol. 1, no. 2, pp. 54–66, Dec. 2021, doi: 10.47709/brilliance.v1i2.1228.

[6] I. Pratiwi, N. Suarna, and T. Suprapti, "IMPLEMENTASI ASSOCIATION RULES MINING UNTUK ANALISIS POLA PEMBELIAN PAKET KUOTA PERDANA PELANGGAN XL MENGGUNAKAN ALGORTIMA APRIORI (STUDI KASUS: PT. XL AXIATA TEGAL)," 2024.

[7] E. Castaño Camps, "Introduction to Time Series and Forecasting," 2022.

[8] Brilliant M, Handoko D, and Sriyanto, "Implementation_of_Data_Mining_Using_Asso," *Implementation of Data Mining Using Association Rules for Transactional Data Analysis*, pp. 177–180, 2017.

[9] D. Mutiara and Eriyanto, "Analisis Framing Pemberitaan Kasus Kekerasan pada Orientasi Pengenalan Kampus," *Jurnal Komunikasi Global*, vol 9, no. 1, 2020.

[10] S. Pahmi, R. Hopipah, D. A. Saputri, T. P. Dewi, H. Yulita, and A. Widowati, "Studi Literatur Terhadap Kekerasan di Kalangan Remaja," *Jurnal Basicedu*, vol. 8, no. 1, pp. 911–920, Dec. 2023, doi: 10.31004/basicedu.v8i1.6354.

[11] Dwi Jasuma, B. (n.d.). news-scraper. GitHub. Retrieved from https://github.com/binsarjr/news-scraper.git

[12] G. Kaur, "Association Rule Mining: A Survey," (IJCSIT) *International Journal of Computer Science and Information Technologies*,vol. 5, no. 2, pp. 2320-2324, 2014.

[13] P. Pillai and D. Amin, "Understanding the requirements Of the Indian IT industry using web scrapping," *Procedia Comput Sci*, vol. 172, pp. 308–313, 2020, doi: https://doi.org/10.1016/j.procs.2020.05.050.

[14] L. J. Tjahyana and F. Lesmana, "Entity Sentiment Analysis with the Netray Monitoring Tool in Indonesian Online News Media on the Fuel Price Hike," *Information and Media*, vol. 99, pp. 106–125, 2024, doi: 10.15388/IM.2024.99.6.

[15] S. J. Blair, Y. Bi, and M. D. Mulvenna, "Aggregated topic models for increasing social media topic coherence," *Applied Intelligence*, vol. 50, no. 1, pp. 138–156, 2020, doi: 10.1007/s10489-019-01438-z.

[16] Barus, E. S., Zarlis, M., Nasution, Z., & Sutarman, S. (2025). Development of machine learning for forecasting optimization implemented in morphology plant growth. Eastern-European Journal of Enterprise Technologies, 3(2 (135), 42–53. https://doi.org/10.15587/1729-4061.2025.331745

[17] E. S. Barus, M. Zarlis, Z. Nasution and Sutarman, "Forcasting Plant Growth Using Neural Network Time Series," 2019 International Conference of Computer Science and Information Technology (ICoSNIKOM), Medan, Indonesia, 2019, pp. 1-4, doi: 10.1109/ICoSNIKOM48755.2019.9111503

Submitted : 20-07-2025 | Reviewed : 25-08-2025 | Accepted : 02-08-2025