

## PENGEMBANGAN PLATFORM ANALISIS ISU PUBLIK PADA MEDIA SOSIAL INSTAGRAM MENGGUNAKAN PENDEKATAN TOPIC MODELING DAN TEXT MINING

Nuzul Hikmah, Dwi Putri Kartini, Kurnia Iswardani, Abd Wahid

Universitas Panca Marga

n.hikmah1807@upm.ac.id

### ABSTRAK

Media sosial, khususnya Instagram, menghasilkan data teks dalam jumlah besar yang berpotensi dimanfaatkan untuk mengidentifikasi isu publik yang berkembang di masyarakat. Namun, proses analisis data media sosial umumnya masih dilakukan menggunakan beberapa perangkat lunak yang terpisah sehingga kurang efisien dan memerlukan kemampuan teknis yang relatif tinggi. Penelitian ini bertujuan mengembangkan Platform Analisis Isu Publik pada Media Sosial Instagram menggunakan pendekatan Topic Modeling dan Text Mining berbasis web untuk mengintegrasikan seluruh tahapan analisis dalam satu aplikasi. Platform dikembangkan menggunakan bahasa pemrograman Python dengan framework Streamlit serta menerapkan metode Latent Dirichlet Allocation (LDA) untuk proses topic modeling. Fitur utama platform meliputi crawling data Instagram berdasarkan hashtag, manajemen dataset, preprocessing teks, penentuan jumlah topik optimal menggunakan Coherence Score, pemodelan topik, dan visualisasi hasil analisis. Hasil implementasi menunjukkan bahwa platform mampu mengintegrasikan seluruh tahapan analisis dalam satu lingkungan kerja. Pengujian topic modeling pada jumlah topik 2–10 menghasilkan Coherence Score tertinggi sebesar 0,5237 pada konfigurasi 5 topik, yang kemudian digunakan sebagai model akhir untuk identifikasi topik dominan. Hasil visualisasi menunjukkan adanya beberapa kelompok isu publik yang berkaitan dengan ekonomi dan harga, pendidikan, bantuan sosial dan kesehatan, kemiskinan, serta lingkungan. Platform yang dikembangkan dapat mendukung proses analisis isu publik secara lebih sistematis dan terintegrasi serta menjadi dasar pengembangan analisis prioritas isu publik berbasis data.

**Kata kunci :** *Instagram, Latent Dirichlet Allocation, platform analisis, text mining, topic modeling.*

### 1. PENDAHULUAN

Perkembangan media sosial telah mengubah cara masyarakat menyampaikan informasi, opini, maupun tanggapan terhadap berbagai isu publik. Berbagai permasalahan yang berkaitan dengan pendidikan, kesehatan, ekonomi, lingkungan, maupun kebijakan pemerintah banyak dibahas melalui unggahan dan interaksi pengguna media sosial. Informasi tersebut tersimpan dalam bentuk data teks yang terus bertambah setiap waktu sehingga berpotensi dimanfaatkan sebagai sumber informasi untuk memahami isu yang berkembang di masyarakat. Oleh karena itu, media sosial tidak hanya berfungsi sebagai sarana komunikasi, tetapi juga menjadi sumber data yang penting dalam mendukung proses analisis dan pengambilan keputusan berbasis data[1],[2].

Instagram merupakan salah satu platform media sosial yang banyak digunakan oleh masyarakat Indonesia untuk berbagi informasi melalui caption, hashtag, komentar, maupun berbagai bentuk interaksi lainnya. Aktivitas tersebut menghasilkan data dalam jumlah besar dengan karakteristik yang tidak terstruktur sehingga sulit dianalisis secara manual. Kondisi tersebut mendorong pemanfaatan teknik text mining untuk mengekstraksi informasi penting dari kumpulan data teks secara otomatis sehingga pola pembahasan dan isu yang berkembang dapat diidentifikasi secara lebih efektif[1],[3].

Salah satu pendekatan yang banyak digunakan dalam text mining adalah topic modeling, yaitu teknik yang mampu menemukan kelompok topik tersembunyi berdasarkan pola kemunculan kata pada sekumpulan dokumen. Berbagai penelitian menunjukkan bahwa topic modeling mampu membantu proses identifikasi isu dominan pada data media sosial tanpa memerlukan proses pelabelan data secara manual. Pendekatan ini telah diterapkan pada berbagai platform media sosial untuk menganalisis opini publik, mengevaluasi kebijakan pemerintah, maupun memahami kecenderungan pembahasan masyarakat terhadap suatu peristiwa [4],[3],[2].

Metode Latent Dirichlet Allocation (LDA) merupakan salah satu metode topic modeling yang paling banyak digunakan karena mampu merepresentasikan setiap dokumen sebagai kombinasi beberapa topik dan menghasilkan interpretasi topik yang relatif baik. Untuk memperoleh model yang optimal, kualitas topik umumnya dievaluasi menggunakan Coherence Score, sehingga jumlah topik yang digunakan tidak hanya ditentukan berdasarkan percobaan, tetapi juga didukung oleh ukuran evaluasi yang objektif[5], [6],[7].

Beberapa penelitian terdahulu telah memanfaatkan LDA dalam analisis data media sosial. Buana et al. mengembangkan Social Media Analytics Tool berbasis Instagram menggunakan

topic modeling untuk mendukung strategi digital marketing[1]. Choirinnisa et al. menerapkan LDA untuk menganalisis opini publik terhadap Kementerian Keuangan pada platform X (Twitter) melalui tahapan preprocessing, perhitungan coherence score, dan visualisasi hasil analisis[4]. Penelitian lain juga menunjukkan bahwa topic modeling mampu mengidentifikasi pola pembahasan masyarakat terhadap berbagai isu publik pada media sosial secara efektif[3],[2].

Meskipun demikian, sebagian besar penelitian sebelumnya masih berfokus pada penerapan metode topic modeling dan interpretasi hasil analisis. Proses pengumpulan data, pengelolaan dataset, preprocessing, evaluasi jumlah topik, pemodelan topik, hingga visualisasi hasil umumnya masih dilakukan menggunakan beberapa perangkat lunak yang berbeda. Pendekatan tersebut menyebabkan proses analisis menjadi kurang efisien, sulit direplikasi, serta memerlukan kemampuan teknis yang relatif tinggi dari pengguna[1],[8],[2].

Berdasarkan kesenjangan tersebut, penelitian ini memberikan kontribusi berupa pengembangan platform analisis isu publik berbasis web yang mengintegrasikan keseluruhan tahapan analisis data Instagram dalam satu lingkungan kerja. Kebaruan penelitian ini terletak pada integrasi proses crawling data Instagram, manajemen dataset, preprocessing teks, penentuan jumlah topik optimal menggunakan Coherence Score, pemodelan topik menggunakan Latent Dirichlet Allocation (LDA), serta visualisasi hasil analisis dalam satu platform. Dengan demikian, penelitian ini tidak hanya menerapkan LDA untuk mengidentifikasi topik, tetapi juga menyediakan alur analisis yang terintegrasi mulai dari pengumpulan data hingga penyajian hasil yang dapat digunakan untuk membantu interpretasi isu publik.

Penelitian ini bertujuan mengembangkan Platform Analisis Isu Publik pada Media Sosial Instagram Menggunakan Pendekatan Topic Modeling dan Text Mining berbasis web. Platform yang dikembangkan diharapkan dapat mendukung proses pengumpulan dan pengolahan data Instagram serta membantu pengguna melakukan identifikasi dan interpretasi isu publik secara lebih sistematis, efektif, dan efisien tanpa harus menggunakan beberapa perangkat lunak secara terpisah.

## 2. TINJAUAN PUSTAKA

### 2.1. Text Mining

Text mining merupakan pendekatan yang digunakan untuk mengekstraksi informasi dan pengetahuan dari data teks yang tidak terstruktur melalui pemanfaatan teknik natural language processing (NLP), statistik, dan machine learning[9]. Pendekatan ini memungkinkan data teks dalam jumlah besar diolah menjadi informasi yang lebih terstruktur sehingga dapat dimanfaatkan untuk berbagai kebutuhan analisis, seperti identifikasi

topik, analisis opini, deteksi tren, hingga pengambilan keputusan berbasis data[2],[10].

Perkembangan media sosial telah meningkatkan pemanfaatan text mining sebagai salah satu metode untuk menganalisis perilaku pengguna dan isu-isu yang berkembang di masyarakat. Data yang berasal dari media sosial umumnya memiliki karakteristik berukuran besar, dinamis, serta tidak terstruktur sehingga memerlukan tahapan pengolahan sebelum dilakukan analisis. Oleh karena itu, proses preprocessing menjadi tahapan penting dalam text mining untuk meningkatkan kualitas data dan akurasi hasil analisis[1],[11],[12].

Pada penelitian ini, proses preprocessing meliputi case folding, tokenisasi, pembersihan URL, mention, emoji, angka, karakter khusus, serta stopword removal. Tahapan tersebut bertujuan mengurangi noise pada data sehingga informasi yang digunakan pada proses pemodelan topik lebih representatif terhadap isi dokumen[4],[13].

### 2.2. Topic Modeling

Topic modeling merupakan salah satu teknik dalam text mining yang bertujuan menemukan struktur topik tersembunyi (latent topics) dari sekumpulan dokumen tanpa memerlukan proses pelabelan data secara manual. Teknik ini bekerja dengan mengidentifikasi pola kemunculan kata yang sering muncul secara bersamaan sehingga membentuk kelompok topik yang merepresentasikan suatu pembahasan tertentu[8],[10].

Dalam beberapa tahun terakhir, topic modeling telah banyak diterapkan pada data media sosial untuk mengidentifikasi isu publik, menganalisis opini masyarakat, serta memahami kecenderungan pembahasan terhadap suatu peristiwa. Berbagai penelitian menunjukkan bahwa topic modeling mampu membantu proses eksplorasi informasi dari data teks berukuran besar secara lebih efektif dibandingkan analisis manual[3],[14].

Meskipun berbagai metode topic modeling telah dikembangkan, seperti Non-negative Matrix Factorization (NMF), Top2Vec, maupun BERTopic, metode Latent Dirichlet Allocation (LDA) masih menjadi salah satu pendekatan yang paling banyak digunakan karena mampu menghasilkan representasi topik yang mudah diinterpretasikan dan memiliki performa yang stabil pada berbagai domain penelitian[8],[10].

### 2.3. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) merupakan metode topic modeling berbasis probabilistik yang diperkenalkan oleh Blei et al.[6]. Metode ini mengasumsikan bahwa setiap dokumen terdiri atas beberapa topik dengan proporsi tertentu, sedangkan setiap topik direpresentasikan oleh distribusi probabilitas dari sejumlah kata. Berdasarkan asumsi tersebut, LDA mampu mengidentifikasi hubungan

antara dokumen dan topik sehingga menghasilkan representasi topik yang dapat digunakan untuk memahami isi kumpulan dokumen secara lebih sistematis.

Secara generatif, setiap dokumen  $d$  memiliki distribusi topik  $\theta_d$  yang dihasilkan dari distribusi Dirichlet dengan parameter  $\alpha$ , sedangkan setiap topik  $k$  memiliki distribusi kata  $\phi_k$  yang dihasilkan dari distribusi Dirichlet dengan parameter  $\beta$ . Untuk setiap posisi kata  $n$ , LDA terlebih dahulu menentukan topik laten  $z_{d,n}$  berdasarkan distribusi topik dokumen, kemudian menentukan kata  $w_{d,n}$  berdasarkan distribusi kata pada topik tersebut. Proses tersebut dapat dinyatakan sebagai:

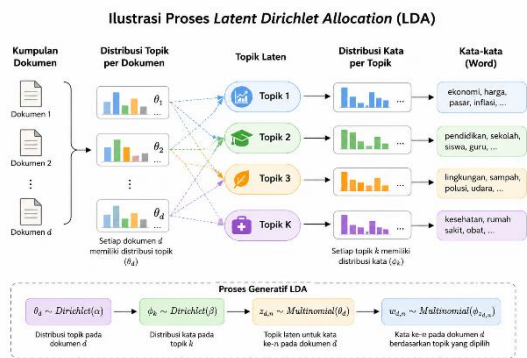
$$\theta_d \sim \text{Dirichlet}(\alpha)$$

$$\phi_k \sim \text{Dirichlet}(\beta)$$

$$z_{d,n} \sim \text{Multinomial}(\theta_d)$$

$$w_{d,n} \sim \text{Multinomial}(\phi_{z_{d,n}})$$

Dengan  $\theta_d$  merupakan distribusi topik pada dokumen  $d$ ,  $\phi_k$  merupakan distribusi kata pada topik  $k$ ,  $z_{d,n}$  merupakan topik laten dari kata ke-  $n$  pada dokumen  $d$ , dan  $w_{d,n}$  merupakan kata yang diamati. Parameter  $\alpha$  dan  $\beta$  merupakan parameter Dirichlet yang mengatur distribusi topik pada dokumen dan distribusi kata pada topik.



Gambar 1. Ilustrasi proses Latent Dirichlet Allocation (LDA)

Ilustrasi pada Gambar 1. menunjukkan bahwa LDA menghubungkan dokumen dengan topik laten melalui distribusi probabilitas. Satu dokumen dapat memiliki lebih dari satu topik dengan proporsi yang berbeda, sementara setiap topik dicirikan oleh sekumpulan kata dengan probabilitas tertentu. Hasil

pemodelan tersebut selanjutnya dapat digunakan untuk menentukan topik dominan pada setiap dokumen.

Pada implementasinya, kualitas model LDA dipengaruhi oleh beberapa parameter, salah satunya adalah jumlah topik (number of topics). Penentuan jumlah topik yang kurang sesuai dapat menyebabkan topik yang dihasilkan sulit diinterpretasikan atau memiliki tingkat kemiripan yang tinggi. Oleh karena itu, penelitian ini menggunakan Coherence Score sebagai ukuran evaluasi untuk menentukan jumlah topik yang paling optimal[7].

Coherence Score mengukur tingkat keterkaitan semantik antar kata pada setiap topik yang dihasilkan. Semakin tinggi nilai Coherence Score, semakin baik kualitas topik yang terbentuk sehingga hasil topic modeling menjadi lebih mudah dipahami dan diinterpretasikan oleh pengguna[7],[13].

Coherence Score digunakan untuk mengevaluasi kualitas topik yang dihasilkan oleh model LDA berdasarkan tingkat keterkaitan antarkata dalam suatu topik. Penelitian ini menggunakan C\_v coherence, salah satu ukuran *topic coherence* yang mengevaluasi keterkaitan kata-kata dalam topik berdasarkan hubungan kemunculan kata pada konteks dokumen. Nilai C\_v yang lebih tinggi menunjukkan keterkaitan antarkata yang lebih baik sehingga topik yang dihasilkan cenderung lebih mudah diinterpretasikan [7].

Dalam penelitian ini, C\_v coherence digunakan untuk membandingkan beberapa alternatif jumlah topik pada model LDA. Model dengan nilai C\_v tertinggi dipilih sebagai konfigurasi jumlah topik yang digunakan dalam proses analisis selanjutnya. Pendekatan ini memungkinkan penentuan jumlah topik dilakukan berdasarkan ukuran evaluasi yang terukur, bukan hanya berdasarkan pertimbangan subjektif.

## 2.4. Penelitian Terdahulu

Penelitian mengenai penerapan topic modeling pada media sosial telah banyak dilakukan dengan berbagai tujuan dan objek penelitian. Namun demikian, sebagian besar penelitian masih berfokus pada proses analisis data menggunakan metode tertentu tanpa mengembangkan platform yang mengintegrasikan seluruh tahapan analisis.

Tabel 1. Perbandingan penelitian terdahulu

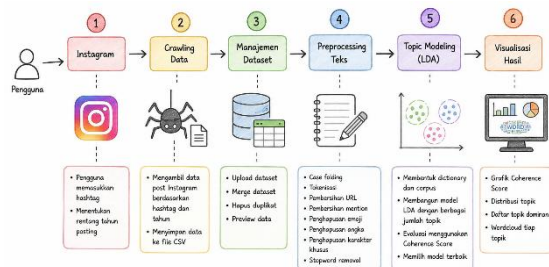
| Peneliti                      | Metode         | Kontribusi                                                                                        | Keterbatasan                                                                                                     |
|-------------------------------|----------------|---------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|
| Buana <i>et al.</i> [1]       | Topic Modeling | Mengembangkan <i>social media analytics</i> berbasis Instagram untuk mendukung digital marketing. | Belum berfokus pada analisis isu publik dan belum mengintegrasikan seluruh tahapan analisis dalam satu platform. |
| Choirinnisa <i>et al.</i> [4] | LDA            | Menganalisis opini publik pada media sosial X menggunakan LDA dan <i>Coherence Score</i> .        | Berfokus pada hasil topic modeling tanpa pengembangan platform analisis.                                         |

| Peneliti                    | Metode                       | Kontribusi                                                                                                                                                                       | Keterbatasan                                                                                    |
|-----------------------------|------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| Suhaeni <i>et al.</i> [4]   | LDA                          | Mengidentifikasi topik dominan pada isu publik menggunakan LDA.                                                                                                                  | Implementasi terbatas pada analisis topik dan belum menyediakan fitur pengelolaan data.         |
| Fahlevvi <i>et al.</i> [13] | LDA                          | Menerapkan LDA pada data portal berita untuk memperoleh topik utama.                                                                                                             | Analisis hanya dilakukan pada data yang telah tersedia tanpa integrasi proses pengumpulan data. |
| Maulana <i>et al.</i> [11]  | Sentiment Analysis & LDA     | Mengombinasikan analisis sentimen dan topic modeling pada media sosial.                                                                                                          | Tidak mengembangkan platform yang mengintegrasikan proses analisis secara menyeluruh.           |
| <b>Penelitian ini</b>       | <b>Text Mining &amp; LDA</b> | <b>Mengembangkan platform analisis isu publik berbasis web yang mengintegrasikan crawling, manajemen dataset, preprocessing, topic modeling, dan visualisasi hasil analisis.</b> | -                                                                                               |

Berdasarkan penelitian-penelitian tersebut dapat diketahui bahwa pemanfaatan topic modeling menggunakan LDA telah banyak diterapkan pada berbagai sumber data, khususnya media sosial. Akan tetapi, sebagian besar penelitian masih berorientasi pada penerapan metode analisis tanpa mengintegrasikan proses pengumpulan data, pengelolaan dataset, preprocessing, pemodelan topik, dan visualisasi hasil dalam satu platform. Oleh karena itu, penelitian ini mengembangkan Platform Analisis Isu Publik pada Media Sosial Instagram yang mengintegrasikan seluruh tahapan tersebut sehingga proses analisis dapat dilakukan secara lebih efektif, efisien, dan mudah digunakan oleh pengguna.

### 3. METODE PENELITIAN

Penelitian ini mengembangkan sebuah platform berbasis web untuk mendukung proses analisis isu publik pada media sosial Instagram menggunakan pendekatan *text mining* dan *topic modeling*. Platform dikembangkan menggunakan bahasa pemrograman Python dengan framework Streamlit yang memungkinkan integrasi seluruh tahapan analisis dalam satu lingkungan kerja. Secara umum, tahapan penelitian meliputi proses crawling data, manajemen dataset, preprocessing teks, pemodelan topik menggunakan Latent Dirichlet Allocation (LDA), serta visualisasi hasil analisis.



Gambar 2. Tahapan penelitian

#### 3.1. Crawling Data Instagram

Dataset penelitian diperoleh dari media sosial Instagram melalui proses *crawling* berdasarkan *hashtag* yang berkaitan dengan isu publik. Pengumpulan data dilakukan dengan menentukan

*hashtag*, rentang tahun posting, dan jumlah *scroll* sebagai parameter proses *crawling*. Data yang diperoleh berupa posting Instagram yang memenuhi kriteria pencarian berdasarkan *hashtag* dan rentang waktu yang ditentukan.

Dataset hasil *crawling* memuat beberapa atribut utama, yaitu *caption*, jumlah *likes*, jumlah *comments*, tanggal posting, dan tanggal *crawling*. Atribut *caption* digunakan sebagai data teks utama dalam proses *preprocessing* dan *topic modeling*, sedangkan jumlah *likes*, jumlah *comments*, dan tanggal posting digunakan sebagai informasi pendukung untuk menggambarkan karakteristik posting. Tanggal *crawling* digunakan untuk mencatat waktu ketika data dikumpulkan.

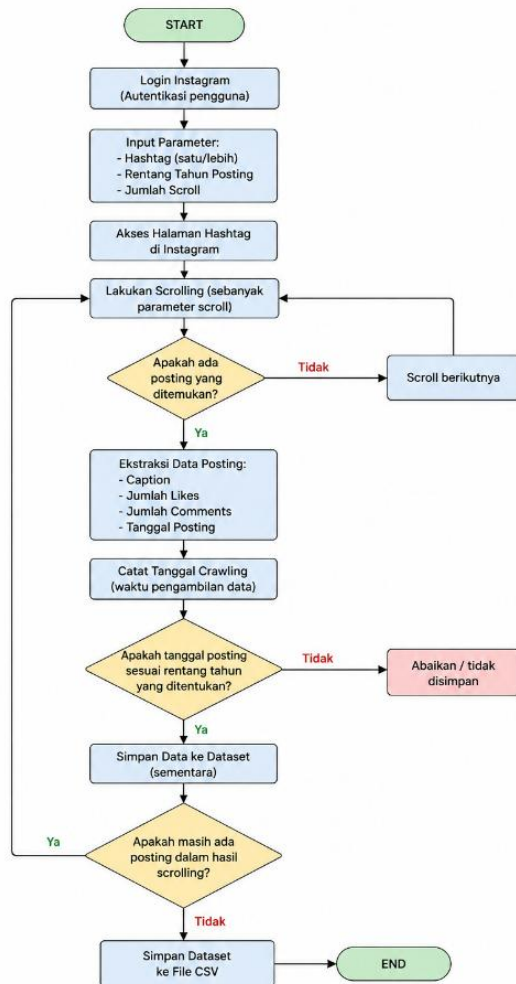
Tabel 2. Atribut dataset hasil crawling

| No. | Atribut          | Keterangan                                | Penggunaan                                           |
|-----|------------------|-------------------------------------------|------------------------------------------------------|
| 1   | Caption          | Teks yang terdapat pada posting Instagram | Input <i>preprocessing</i> dan <i>topic modeling</i> |
| 2   | Likes            | Jumlah tanda suka pada posting            | Informasi pendukung                                  |
| 3   | Comments         | Jumlah komentar pada posting              | Informasi pendukung                                  |
| 4   | Tanggal posting  | Waktu publikasi posting                   | Informasi pendukung dan penyaringan periode          |
| 5   | Tanggal crawling | Waktu pengambilan data                    | Dokumentasi pengumpulan data                         |

Proses *crawling* dilakukan melalui browser setelah pengguna melakukan proses login pada akun Instagram. Setelah autentikasi berhasil, pengguna memasukkan satu atau beberapa *hashtag*, rentang tahun posting, dan jumlah *scroll*. Sistem kemudian mengakses halaman *hashtag* yang ditentukan dan melakukan *scrolling* secara bertahap untuk memperoleh posting yang tersedia.

Pada setiap posting yang berhasil ditemukan, sistem mengambil informasi berupa *caption*, jumlah *likes*, jumlah *comments*, dan tanggal posting. Sistem

juga mencatat tanggal ketika proses *crawling* dilakukan. Data yang diperoleh kemudian diseleksi berdasarkan rentang tahun yang telah ditentukan pengguna. Posting yang memenuhi kriteria dimasukkan ke dalam dataset, sedangkan data yang berada di luar rentang yang ditentukan tidak digunakan.



Gambar 3. Flowchart Mekanisme Crawling Data Instagram

Data hasil *crawling* disimpan dalam format CSV sehingga dapat digunakan kembali pada proses analisis berikutnya tanpa perlu melakukan pengambilan data ulang. Pemisahan proses pengumpulan data dan analisis memberikan fleksibilitas kepada pengguna dalam mengelola dataset yang akan digunakan.

### 3.2. Manajemen Dataset

Tahap berikutnya adalah manajemen dataset. Modul ini menyediakan fasilitas untuk mengunggah dataset hasil *crawling*, menggabungkan beberapa dataset, menghapus data duplikat, serta menampilkan *preview* data sebelum dilakukan proses analisis.

Keberadaan modul manajemen dataset bertujuan untuk memastikan bahwa data yang digunakan pada tahap berikutnya memiliki kualitas yang baik dan bebas dari duplikasi yang dapat memengaruhi hasil analisis.

### 3.3. Preprocessing Teks

Preprocessing dilakukan untuk membersihkan data teks sebelum proses topic modeling. Tahapan ini bertujuan menghilangkan berbagai komponen yang tidak memberikan kontribusi terhadap proses analisis sehingga kualitas data menjadi lebih baik.

Tahapan preprocessing yang diterapkan seperti yang terlihat pada Gambar 2 berikut:



Gambar 4. Tahapan preprocessing

Hasil preprocessing berupa dataset bersih yang selanjutnya digunakan sebagai masukan pada proses pemodelan topik.

### 3.4. Topic Modeling Menggunakan LDA

Dataset hasil preprocessing digunakan sebagai masukan dalam proses *\*topic modeling\** menggunakan metode Latent Dirichlet Allocation (LDA). Sebelum model dibangun, sistem membentuk *dictionary* dan *corpus* berdasarkan hasil preprocessing.

Untuk memperoleh model yang memiliki kualitas terbaik, dilakukan pengujian beberapa alternatif jumlah topik. Setiap model dievaluasi menggunakan  $C_v$  coherence. Nilai  $C_v$  diperoleh untuk setiap alternatif jumlah topik, kemudian model dengan nilai  $C_v$  tertinggi dipilih sebagai model terbaik dan digunakan pada tahap analisis selanjutnya. Dengan demikian, pemilihan jumlah topik dilakukan berdasarkan ukuran koherensi yang terukur.

Tahapan tersebut memungkinkan sistem menentukan jumlah topik secara lebih objektif sehingga hasil pemodelan lebih mudah diinterpretasikan.

Dalam implementasinya, model LDA dibangun menggunakan sejumlah parameter yang ditetapkan pada proses pelatihan. Parameter yang digunakan meliputi jumlah topik (*num\_topics*), jumlah *passes*, jumlah *iterations*, *random state*, serta parameter distribusi prior *alpha* dan *eta*. Parameter tersebut digunakan secara konsisten pada setiap eksperimen, sedangkan jumlah topik divariasikan untuk memperoleh konfigurasi model yang memberikan kualitas topik terbaik.

Tabel 3. Parameter model LDA

| Parameter           | Nilai             | Keterangan                              |
|---------------------|-------------------|-----------------------------------------|
| <i>num_topics</i>   | 2–10              | Jumlah topik yang diuji                 |
| <i>passes</i>       | [isi sesuai kode] | Jumlah proses pelatihan terhadap corpus |
| <i>iterations</i>   | [isi sesuai kode] | Jumlah iterasi inferensi                |
| <i>random_state</i> | [isi sesuai kode] | Seed untuk menjaga konsistensi hasil    |
| <i>alpha</i>        | [isi sesuai kode] | Prior distribusi topik pada dokumen     |
| <i>eta</i>          | [isi sesuai kode] | Prior distribusi kata pada topik        |

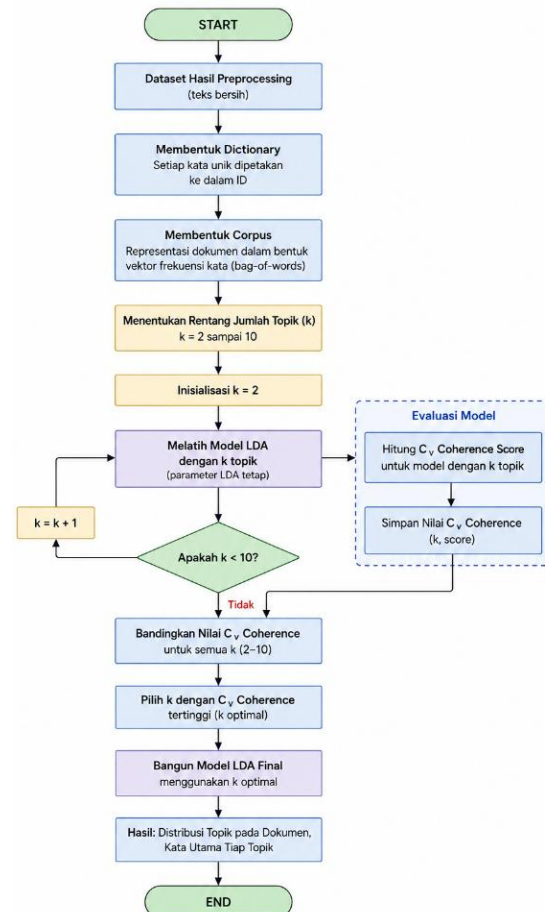
Jumlah topik diuji pada rentang 2 hingga 10. Untuk setiap nilai jumlah topik, model LDA dilatih menggunakan parameter lainnya yang sama. Setiap model kemudian dievaluasi menggunakan  $C_v$  coherence. Model dengan nilai  $C_v$  coherence tertinggi dipilih sebagai model akhir.

### 3.5. Evaluasi Topic Modeling Menggunakan $C_v$ Coherence

Evaluasi dilakukan untuk menentukan konfigurasi jumlah topik yang menghasilkan topik dengan tingkat keterkaitan kata yang paling baik. Penelitian ini menggunakan  $C_v$  coherence sebagai ukuran evaluasi.

$C_v$  coherence merupakan salah satu ukuran topic coherence yang mengevaluasi keterkaitan kata-kata dalam suatu topik berdasarkan hubungan kemunculan kata pada konteks dokumen. Nilai  $C_v$  coherence yang lebih tinggi menunjukkan keterkaitan antarkata yang lebih baik sehingga topik yang dihasilkan cenderung lebih mudah diinterpretasikan.

Pada penelitian ini,  $C_v$  coherence dihitung untuk setiap model LDA dengan jumlah topik 2 hingga 10. Nilai yang diperoleh kemudian dibandingkan untuk menentukan konfigurasi jumlah topik yang paling optimal. Model dengan nilai  $C_v$  coherence tertinggi digunakan sebagai model LDA akhir untuk proses identifikasi topik dominan.



Gambar 5. Flowchart proses pemodelan dan evaluasi LDA

### 3.6. Visualisasi Hasil Analisis

Model LDA terbaik digunakan untuk menghasilkan distribusi topik pada setiap dokumen. Setiap posting diberikan probabilitas terhadap beberapa topik, kemudian topik dengan probabilitas tertinggi digunakan sebagai topik dominan pada posting tersebut.

Hasil analisis divisualisasikan dalam beberapa bentuk, yaitu grafik  $C_v$  coherence, distribusi topik, daftar kata utama pada setiap topik, dan *wordcloud*. Grafik  $C_v$  coherence digunakan untuk menunjukkan perbandingan kualitas model pada berbagai jumlah topik. Distribusi topik digunakan untuk melihat jumlah posting yang termasuk ke dalam masing-masing topik.

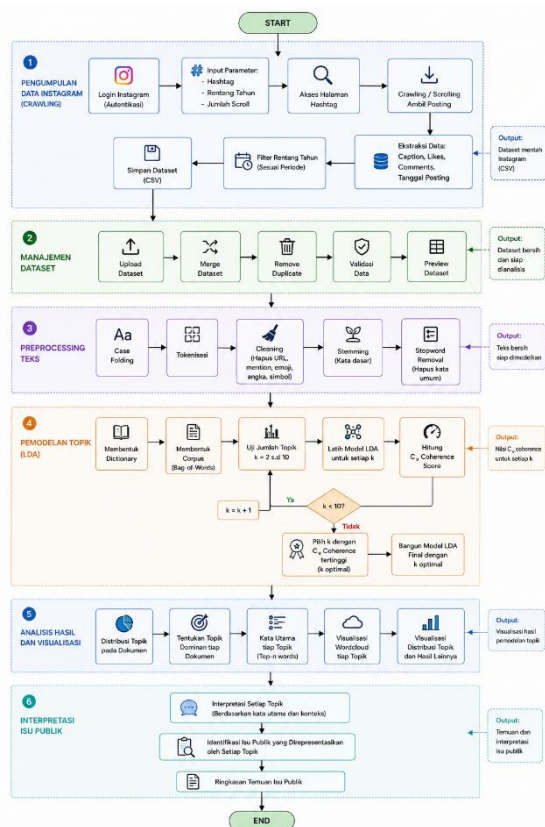
Selanjutnya, daftar kata utama dan *wordcloud* digunakan untuk membantu proses interpretasi topik. Kata-kata dengan probabilitas atau frekuensi tinggi pada suatu topik digunakan sebagai dasar

untuk memberikan label dan memahami isu publik yang direpresentasikan oleh topik tersebut.

Interpretasi isu publik dilakukan dengan mempertimbangkan keterkaitan beberapa kata utama dalam satu topik, bukan berdasarkan satu kata secara terpisah. Dengan demikian, hasil LDA digunakan sebagai dasar untuk mengidentifikasi kecenderungan pembahasan isu publik dalam dataset Instagram.

### 3.7. Alur Algoritma Penelitian

Secara keseluruhan, algoritma penelitian dapat diringkas sebagai berikut:



Gambar 6. Flowchart algoritma analisis isu publik

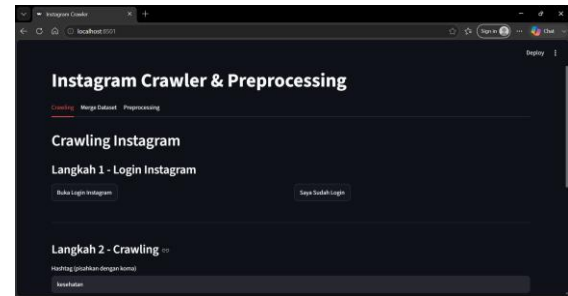
## 4. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan platform berbasis web untuk mendukung proses analisis isu publik pada media sosial Instagram menggunakan pendekatan text mining dan topic modeling. Platform dikembangkan menggunakan bahasa pemrograman Python dengan framework Streamlit dan mengintegrasikan tahapan pengumpulan data, manajemen dataset, preprocessing, pemodelan topik menggunakan Latent Dirichlet Allocation (LDA), evaluasi model, serta visualisasi hasil analisis.

Hasil penelitian dibahas berdasarkan implementasi setiap modul, hasil preprocessing, evaluasi model LDA, hasil topic modeling, interpretasi isu publik, serta pengujian terhadap fungsionalitas platform.

### 4.1. Implementasi Crawling Data Instagram

Modul *crawling* merupakan tahap awal dalam proses analisis yang berfungsi mengumpulkan data posting Instagram berdasarkan hashtag yang dimasukkan oleh pengguna. Pada modul ini, pengguna menentukan hashtag yang akan dianalisis, jumlah scroll, serta rentang tahun posting yang akan diambil. Informasi tersebut digunakan sistem sebagai parameter dalam proses pengambilan data.

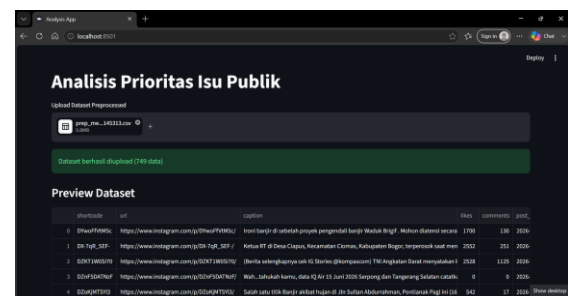


Gambar 7. Antarmuka modul crawling data instagram

Berdasarkan Gambar 3, proses crawling diawali dengan login ke akun Instagram melalui browser. Setelah proses login berhasil dilakukan, pengguna memasukkan hashtag, jumlah scroll, serta rentang tahun posting yang diinginkan. Selanjutnya sistem akan melakukan proses crawling secara otomatis dan menyimpan hasilnya dalam bentuk dataset berformat CSV. Dataset yang dihasilkan memuat informasi berupa caption, jumlah likes, jumlah comments, tanggal posting, dan tanggal crawling. Pemisahan proses crawling dari proses analisis memungkinkan pengguna mengumpulkan data secara bertahap tanpa harus mengulangi proses analisis dari awal.

### 4.2. Implementasi Modul Manajemen Dataset

Setelah proses crawling selesai, dataset dapat dikelola melalui modul manajemen dataset. Modul ini menyediakan fasilitas untuk mengunggah dataset hasil crawling, menggabungkan beberapa dataset, menghapus data duplikat, serta melakukan validasi sebelum analisis dilakukan.



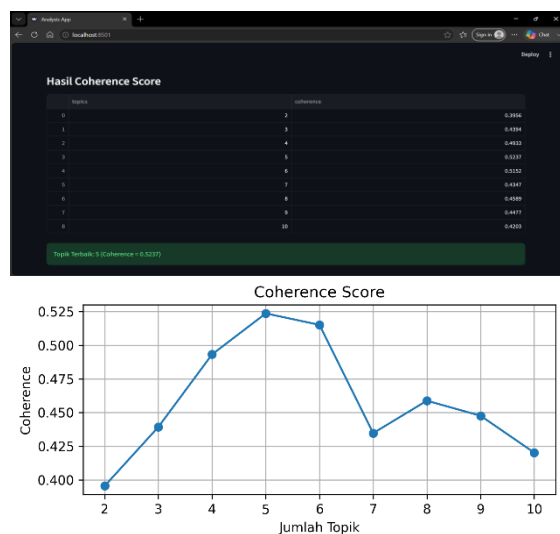
Gambar 8. Antarmuka modul upload dan preview dataset

Gambar 4 menunjukkan proses unggah dataset hasil crawling ke dalam sistem. Setelah dataset

berhasil diunggah, sistem menampilkan jumlah data yang berhasil dibaca serta melakukan validasi struktur data. Preview dataset ditampilkan sebagaimana terlihat pada Gambar 4 sehingga pengguna dapat memastikan bahwa data yang akan dianalisis telah sesuai. Tahapan ini penting untuk mengurangi kemungkinan kesalahan analisis akibat struktur data yang tidak lengkap atau tidak sesuai.

#### 4.3. Implementasi Topic Modeling Menggunakan LDA

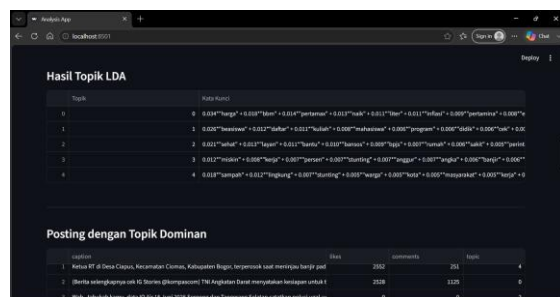
Setelah dataset berhasil diproses, sistem melakukan pembentukan model Latent Dirichlet Allocation (LDA). Sebelum model akhir dibangun, dilakukan pengujian beberapa alternatif jumlah topik untuk memperoleh model yang paling representatif.



Gambar 9. Hasil Perhitungan Coherence Score

Berdasarkan Gambar 6, pengujian dilakukan pada jumlah topik 2 hingga 10. Evaluasi menggunakan  $C_v$  coherence menunjukkan bahwa nilai tertinggi diperoleh pada konfigurasi 5 topik, yaitu sebesar 0,5237. Oleh karena itu, model dengan 5 topik dipilih sebagai model akhir untuk proses identifikasi topik dominan.

Setelah jumlah topik optimal diperoleh, sistem menghasilkan daftar kata kunci utama pada setiap topik.



Gambar 10. Hasil topic modeling menggunakan LDA

Gambar 7 memperlihatkan lima topik yang berhasil diidentifikasi beserta kata-kata yang memiliki probabilitas tertinggi pada masing-masing topik. Kata-kata tersebut digunakan sebagai dasar dalam proses interpretasi topik sesuai dengan konteks isu yang dibahas masyarakat. Selain menghasilkan daftar topik, sistem juga menentukan topik dominan pada setiap posting. Hasil pada Gambar 7 menunjukkan bahwa setiap posting berhasil dikelompokkan ke dalam salah satu topik dominan berdasarkan probabilitas tertinggi yang dihasilkan oleh model LDA.

Kata kunci tersebut diperoleh berdasarkan kata-kata dengan probabilitas tertinggi pada masing-masing topik dan digunakan sebagai dasar dalam memberikan interpretasi terhadap isu publik yang terbentuk.

Tabel 4. Hasil Topik Modeling LDA

| Topik | Kata Kunci Utama                                                                      | Interpretasi Topik                                    |
|-------|---------------------------------------------------------------------------------------|-------------------------------------------------------|
| 0     | harga, BBM, Pertamina, naik, inflasi, ekonomi, rupiah, minyak                         | <b>Ekonomi, BBM, dan Inflasi</b>                      |
| 1     | beasiswa, daftar, kuliah, mahasiswa, program, didik, ajar, buka                       | <b>Pendidikan dan Beasiswa</b>                        |
| 2     | sehat, layan, bantu, bansos, BPJS, rumah, sakit, pemerintah, dokter                   | <b>Kesehatan dan Layanan BPJS</b>                     |
| 3     | miskin, kerja, persen, stunting, anggur, angka, banjir, anak, tumbuh, makan           | <b>Kemiskinan, Ketenagakerjaan, dan Kesejahteraan</b> |
| 4     | sampah, lingkung, stunting, warga, kota, masyarakat, kerja, pemerintah, kelola, sehat | <b>Lingkungan dan Persampahan</b>                     |

Berdasarkan Tabel 4, setiap topik menunjukkan karakteristik kata yang berbeda sehingga dapat digunakan untuk mengidentifikasi beberapa kelompok isu publik yang muncul dalam dataset. Interpretasi masing-masing topik dijelaskan sebagai berikut.

#### Topik 0 – Ekonomi, BBM, dan Inflasi

Topik 0 didominasi oleh kata *harga*, *BBM*, *Pertamax*, *naik*, *inflasi*, *Pertamina*, *ekonomi*, *rupiah*, dan *minyak*. Keterkaitan kata-kata tersebut menunjukkan adanya pembahasan publik mengenai kondisi ekonomi, khususnya perubahan harga bahan bakar minyak dan kaitannya dengan inflasi. Kemunculan kata *Pertamax*, *Pertamina*, dan *minyak* memperlihatkan bahwa pembahasan ekonomi

dalam topik ini memiliki keterkaitan dengan isu energi dan harga BBM.

Dari perspektif isu publik, topik ini menunjukkan perhatian masyarakat terhadap perubahan harga dan kondisi ekonomi yang berpotensi berkaitan dengan pengeluaran masyarakat. Dengan demikian, isu ekonomi yang teridentifikasi tidak hanya bersifat makro seperti inflasi dan nilai rupiah, tetapi juga berkaitan dengan komoditas yang secara langsung dirasakan masyarakat, yaitu bahan bakar.

#### Topik 1 – Pendidikan dan Beasiswa

Topik 1 ditandai oleh kata *beasiswa*, *daftar*, *kuliah*, *mahasiswa*, *program*, *didik*, *ajar*, dan *buka*. Kumpulan kata tersebut menunjukkan adanya pembahasan mengenai pendidikan, terutama akses mahasiswa terhadap program pendidikan dan beasiswa.

Kemunculan kata *daftar*, *buka*, dan *beasiswa* menunjukkan bahwa perhatian publik pada topik ini tidak hanya berkaitan dengan pendidikan secara umum, tetapi juga dengan informasi mengenai kesempatan memperoleh dukungan pendidikan. Oleh karena itu, topik ini dapat diinterpretasikan sebagai isu publik terkait akses pendidikan dan dukungan pembiayaan bagi mahasiswa.

#### Topik 2 – Kesehatan dan Layanan BPJS

Topik 2 terdiri atas kata *sehat*, *layan*, *bantu*, *bansos*, *BPJS*, *rumah*, *sakit*, *pemerintah*, dan *dokter*. Pola kata tersebut menunjukkan adanya pembahasan mengenai layanan kesehatan dan dukungan pemerintah terhadap masyarakat.

Kemunculan kata *BPJS*, *rumah sakit*, dan *dokter* secara khusus mengarah pada aspek pelayanan kesehatan, sedangkan kata *bantu* dan *bansos* menunjukkan adanya keterkaitan dengan dukungan sosial. Berdasarkan pola tersebut, isu publik yang direpresentasikan oleh topik ini dapat dipahami sebagai akses dan pelayanan kesehatan serta dukungan pemerintah kepada masyarakat.

#### Topik 3 – Kemiskinan, Ketenagakerjaan, dan Kesejahteraan

Topik 3 didominasi oleh kata *miskin*, *kerja*, *persen*, *stunting*, *angka*, *banjir*, *anak*, *tumbuh*, dan *makan*. Kemunculan kata *miskin*, *kerja*, dan *persen* menunjukkan adanya pembahasan mengenai kondisi sosial-ekonomi dan ketenagakerjaan. Sementara itu, kata *stunting*, *anak*, *tumbuh*, dan *makan* menunjukkan adanya aspek kesejahteraan anak dan pemenuhan kebutuhan dasar.

Topik ini menunjukkan bahwa isu kemiskinan tidak berdiri sendiri, tetapi dapat muncul bersamaan dengan pembahasan mengenai pekerjaan, kondisi anak, dan kesejahteraan masyarakat. Dengan demikian, topik tersebut dapat dipahami sebagai kelompok isu yang

berkaitan dengan kondisi sosial-ekonomi, ketenagakerjaan, dan kesejahteraan masyarakat. Namun, keberadaan kata seperti *banjir* dan *anggur* menunjukkan bahwa topik ini memiliki cakupan kata yang cukup beragam. Hal tersebut menunjukkan salah satu karakteristik model LDA, yaitu satu topik dapat memuat kata-kata yang tidak seluruhnya memiliki keterkaitan substantif yang kuat. Oleh karena itu, interpretasi topik perlu dilakukan berdasarkan pola beberapa kata utama secara bersamaan dan tidak hanya berdasarkan satu kata.

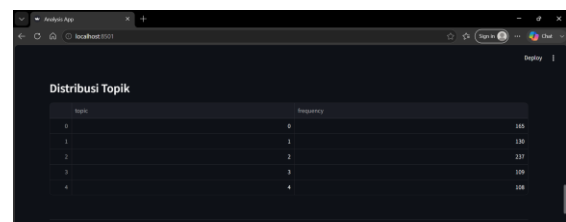
#### Topik 4 – Lingkungan dan Persampahan

Topik 4 ditandai oleh kata *sampah*, *lingkung*, *stunting*, *warga*, *kota*, *masyarakat*, *kerja*, *pemerintah*, *kelola*, dan *sehat*. Kata *sampah*, *lingkung*, *kota*, dan *kelola* menunjukkan pola pembahasan yang kuat mengenai lingkungan, khususnya pengelolaan sampah di wilayah perkotaan. Kemunculan kata *warga*, *masyarakat*, dan *pemerintah* juga menunjukkan adanya keterlibatan masyarakat maupun pemerintah dalam konteks permasalahan tersebut.

Berdasarkan pola tersebut, topik ini dapat diinterpretasikan sebagai isu lingkungan dan pengelolaan persampahan, terutama yang berkaitan dengan pengelolaan lingkungan di tingkat masyarakat dan perkotaan.

#### 4.4. Visualisasi Hasil Analisis

Platform menyediakan beberapa bentuk visualisasi untuk membantu pengguna memahami hasil *topic modeling*, yaitu distribusi topik dan *wordcloud* pada masing-masing topik. Visualisasi tersebut tidak hanya digunakan untuk melihat karakteristik hasil pemodelan, tetapi juga menjadi dasar dalam menginterpretasikan isu publik yang muncul pada data Instagram.



Gambar 11. Distribusi Topik

Distribusi topik pada Gambar 9 menunjukkan jumlah posting yang termasuk ke dalam masing-masing topik hasil pemodelan LDA. Perbedaan jumlah posting pada setiap topik menunjukkan bahwa perhatian masyarakat terhadap isu yang teridentifikasi tidak memiliki intensitas yang sama. Topik dengan jumlah posting lebih besar dapat dipandang sebagai isu yang lebih banyak dibahas dalam dataset penelitian, meskipun tingginya jumlah pembahasan tidak secara langsung menunjukkan tingkat kepentingan atau urgensi suatu isu.

Selanjutnya, sistem membentuk wordcloud untuk setiap topik.



Gambar 12. Wordcloud masing-masing topik

*Wordcloud* pada Gambar 10 digunakan untuk menginterpretasikan karakteristik masing-masing topik berdasarkan kata-kata yang memiliki frekuensi atau bobot relatif tinggi. Hasil visualisasi menunjukkan adanya beberapa kelompok isu publik yang menonjol. Salah satu kelompok memperlihatkan keterkaitan kata ekonomi, harga, dan inflasi, yang mengindikasikan adanya pembahasan mengenai kondisi ekonomi dan perubahan harga sebagai isu yang mendapat perhatian masyarakat. Kelompok kata kuliah, daftar, dan beasiswa menunjukkan isu yang berkaitan dengan akses dan pembiayaan pendidikan. Sementara itu, kemunculan kata bansos, layanan,

dan sehat menunjukkan pembahasan mengenai bantuan sosial dan layanan kesehatan sebagai bagian dari perhatian publik.

Selain itu, terdapat kelompok kata miskin dan persen yang mengarah pada pembahasan mengenai kondisi kemiskinan dan indikator sosial-ekonomi. Kelompok lainnya ditandai dengan kata sampah dan lingkungan, yang menunjukkan adanya perhatian terhadap permasalahan lingkungan, khususnya pengelolaan sampah. Dengan demikian, hasil *topic modeling* tidak hanya menghasilkan kelompok kata secara statistik, tetapi dapat digunakan untuk memetakan beberapa isu publik yang berkembang dalam data Instagram, meliputi ekonomi dan harga, pendidikan, bantuan sosial dan kesehatan, kemiskinan, serta lingkungan.

Temuan tersebut menunjukkan bahwa isu publik yang muncul dalam dataset memiliki karakteristik yang beragam dan mencakup aspek ekonomi maupun sosial. Namun, interpretasi topik tetap perlu dilakukan dengan mempertimbangkan konteks kata-kata yang membentuk setiap topik. LDA mengelompokkan kata berdasarkan pola kemunculannya dalam dokumen, sehingga hasil tersebut menunjukkan kecenderungan pembahasan, bukan hubungan sebab-akibat antarsu. Oleh karena itu, hasil visualisasi lebih tepat digunakan sebagai dasar eksplorasi dan identifikasi awal isu publik yang selanjutnya dapat dianalisis lebih mendalam.

#### 4.5. Pengujian Fungsionalitas Platform

Untuk memastikan bahwa platform yang dikembangkan dapat menjalankan fungsi sesuai kebutuhan, dilakukan pengujian fungsional menggunakan metode Black Box Testing. Pengujian difokuskan pada fungsi utama platform tanpa mempertimbangkan struktur kode internal.

Tabel 5. Hasil Pengujian Black Box

| No. | Fitur            | Skenario Pengujian                        | Hasil yang Diharapkan                 | Hasil    |
|-----|------------------|-------------------------------------------|---------------------------------------|----------|
| 1   | Login Instagram  | Memasukkan akun untuk autentikasi         | Sistem berhasil melakukan autentikasi | Berhasil |
| 2   | Crawling         | Memasukkan hashtag dan parameter crawling | Data berhasil dikumpulkan             | Berhasil |
| 3   | Upload dataset   | Mengunggah file CSV                       | Dataset berhasil dibaca               | Berhasil |
| 4   | Merge dataset    | Menggabungkan beberapa dataset            | Dataset berhasil digabungkan          | Berhasil |
| 5   | Remove duplicate | Menjalankan penghapusan duplikasi         | Data duplikat dihapus                 | Berhasil |
| 6   | Preprocessing    | Menjalankan seluruh tahapan preprocessing | Dataset bersih dihasilkan             | Berhasil |
| 7   | Stemming         | Menjalankan stemming                      | Kata diubah menjadi bentuk dasar      | Berhasil |
| 8   | Topic modeling   | Menjalankan LDA                           | Topik berhasil dibentuk               | Berhasil |
| 9   | C_v coherence    | Menguji jumlah topik 2–10                 | Nilai C_v ditampilkan                 | Berhasil |
| 10  | Visualisasi      | Membuka hasil analisis                    | Grafik dan wordcloud ditampilkan      | Berhasil |
| 11  | Export           | Menyimpan hasil dataset                   | File berhasil disimpan                | Berhasil |

Berdasarkan hasil pengujian, seluruh fungsi yang diuji menghasilkan keluaran sesuai dengan yang diharapkan. Pengujian menunjukkan bahwa fungsi utama platform mulai dari pengumpulan data, pengelolaan dataset, *preprocessing*, pemodelan

LDA, evaluasi  $C_v$  coherence, hingga visualisasi dapat dijalankan melalui platform.

Dengan demikian, pengujian fungsional memberikan bukti bahwa integrasi berbagai tahapan

analisis dalam satu platform dapat berjalan sesuai dengan fungsi yang dirancang.

#### 4.6. Pembahasan

Hasil implementasi menunjukkan bahwa platform yang dikembangkan mampu mengintegrasikan seluruh tahapan analisis isu publik dalam satu aplikasi, mulai dari proses pengumpulan data, pengelolaan dataset, *preprocessing*, pembentukan model LDA, hingga visualisasi hasil analisis. Integrasi tersebut memberikan alur kerja yang lebih sederhana karena pengguna dapat melakukan proses analisis tanpa harus berpindah ke perangkat lunak lain. Selain mendukung proses analisis, pemisahan antara tahap *crawling* dan analisis juga memungkinkan dataset yang telah dikumpulkan untuk digunakan kembali pada proses berikutnya.

Penggunaan  $C_v$  coherence dalam menentukan jumlah topik memberikan dasar kuantitatif dalam pemilihan model. Pada dataset penelitian ini, pengujian dilakukan pada jumlah topik 2 hingga 10 dan menghasilkan nilai  $C_v$  coherence tertinggi sebesar 0,5237 pada lima topik. Peningkatan nilai coherence dari dua hingga lima topik menunjukkan bahwa penambahan jumlah topik pada rentang tersebut masih mampu memberikan pemisahan tema yang lebih baik. Pada jumlah topik yang lebih sedikit, beberapa isu dengan karakteristik berbeda berpotensi tergabung dalam kelompok yang sama sehingga kata-kata yang muncul menjadi lebih umum. Ketika jumlah topik mencapai lima, pembahasan dapat dipisahkan menjadi kelompok isu yang lebih spesifik, yaitu ekonomi dan harga BBM, pendidikan dan beasiswa, kesehatan dan layanan BPJS, kemiskinan dan kesejahteraan, serta lingkungan dan persampahan. Pemisahan tersebut menghasilkan topik yang relatif lebih mudah diinterpretasikan.

Sebaliknya, peningkatan jumlah topik menjadi enam hingga sepuluh tidak menghasilkan nilai coherence yang lebih tinggi. Kondisi tersebut dapat mengindikasikan bahwa pembagian topik yang terlalu banyak menyebabkan informasi dalam dataset menjadi semakin terfragmentasi. Kata-kata yang sebelumnya membentuk kelompok topik yang koheren dapat tersebar ke dalam beberapa kelompok yang lebih kecil sehingga keterkaitan antarkata dalam masing-masing topik menurun. Kondisi ini juga dapat dipengaruhi oleh karakteristik data Instagram yang tidak terstruktur, karena satu kata dapat digunakan dalam beberapa konteks isu yang berbeda. Dengan demikian, penambahan jumlah topik tidak selalu menghasilkan model yang lebih baik. Pada dataset penelitian ini, lima topik memberikan keseimbangan yang lebih baik antara pemisahan tema dan keterkaitan kata dalam topik.

Hasil *topic modeling* menunjukkan bahwa kelima topik memiliki interpretasi yang relatif jelas. Topik 0 yang ditandai oleh kata *harga*, *BBM*,

*Pertamax*, *naik*, *inflasi*, *Pertamina*, *ekonomi*, *rupiah*, dan *minyak* dapat diinterpretasikan sebagai isu ekonomi, BBM, dan inflasi. Kombinasi kata tersebut menunjukkan perhatian terhadap perubahan harga bahan bakar dan kondisi ekonomi yang berkaitan dengan kehidupan masyarakat. Topik 1 yang didominasi oleh kata *beasiswa*, *daftar*, *kuliah*, *mahasiswa*, *program*, *didik*, dan *ajar* mengarah pada isu pendidikan dan beasiswa, khususnya mengenai akses pendidikan dan dukungan pembiayaan bagi mahasiswa.

Topik 2 terdiri atas kata *sehat*, *layan*, *bantu*, *bansos*, *BPJS*, *rumah*, *sakit*, *pemerintah*, dan *dokter*. Pola tersebut mengarah pada isu kesehatan dan layanan BPJS, yang berkaitan dengan akses terhadap pelayanan serta dukungan pemerintah kepada masyarakat. Topik 3 ditandai oleh kata *miskin*, *kerja*, *persen*, *stunting*, *angka*, *banjir*, *anak*, *tumbuh*, dan *makan*. Topik ini merepresentasikan kemiskinan, ketenagakerjaan, dan kesejahteraan, karena kata-kata tersebut berkaitan dengan kondisi sosial-ekonomi, pekerjaan, serta kesejahteraan anak. Sementara itu, Topik 4 yang didominasi oleh kata *sampah*, *lingkung*, *stunting*, *warga*, *kota*, *masyarakat*, *kerja*, *pemerintah*, *kelola*, dan *sehat* mengarah pada isu lingkungan dan persampahan, terutama terkait pengelolaan sampah dan lingkungan di tingkat masyarakat maupun perkotaan.

Meskipun kelima topik dapat diberikan interpretasi yang relatif jelas, hasil pemodelan juga memperlihatkan adanya *overlapping* antar-topik. Beberapa kata seperti *stunting*, *sehat*, *kerja*, dan *pemerintah* muncul pada lebih dari satu topik. Kemunculan *stunting*, misalnya, dapat berkaitan dengan isu kesejahteraan anak pada Topik 3 sekaligus aspek kesehatan masyarakat pada Topik 4. Demikian pula, kata *pemerintah* dapat muncul pada pembahasan kesehatan maupun lingkungan karena kedua isu tersebut berkaitan dengan pelayanan dan kebijakan publik. *Overlapping* tersebut merupakan konsekuensi dari karakteristik LDA yang merepresentasikan topik secara probabilistik, sehingga suatu kata tidak harus hanya berkaitan dengan satu topik. Oleh karena itu, interpretasi hasil tidak dilakukan berdasarkan satu kata secara terpisah, tetapi berdasarkan kombinasi kata-kata utama dan konteks yang membentuk setiap topik.

Secara substantif, hasil tersebut memperlihatkan bahwa isu publik yang muncul dalam dataset Instagram memiliki cakupan yang cukup beragam, mulai dari persoalan ekonomi dan harga BBM hingga pendidikan, kesehatan, kesejahteraan, dan lingkungan. Beberapa isu juga memperlihatkan keterkaitan antardimensi. Isu kesehatan, misalnya, berkaitan dengan bantuan sosial dan layanan pemerintah, sedangkan isu kemiskinan dapat berkaitan dengan pekerjaan, pemenuhan kebutuhan dasar, dan kondisi anak. Hal tersebut memperlihatkan bahwa pembahasan isu

publik di media sosial tidak selalu terbentuk sebagai kategori yang sepenuhnya terpisah, tetapi dapat memiliki dimensi sosial yang saling beririsan.

Pemilihan metode *topic modeling* perlu mempertimbangkan karakteristik data dan kualitas topik yang dihasilkan. Egger dan Yu[10] melakukan perbandingan beberapa metode *topic modeling*, yaitu Latent Dirichlet Allocation (LDA), Non-negative Matrix Factorization (NMF), Top2Vec, dan BERTopic untuk menganalisis data unggahan Twitter. Perbandingan tersebut menunjukkan bahwa masing-masing metode memiliki karakteristik dan kemampuan yang berbeda dalam menghasilkan topik. Temuan tersebut menunjukkan bahwa pemilihan metode *topic modeling* perlu disesuaikan dengan karakteristik data dan tujuan analisis. Dalam penelitian ini, LDA dipilih sebagai metode pemodelan topik dan kualitas model dievaluasi menggunakan *C\_v coherence* untuk menentukan konfigurasi jumlah topik yang paling sesuai dengan dataset Instagram yang digunakan.

Di sisi lain, terdapat keterbatasan yang perlu diperhatikan dalam penggunaan LDA pada data Instagram. LDA bekerja berdasarkan pola distribusi kata sehingga konteks semantik dan hubungan antar kata tidak selalu dapat direpresentasikan secara optimal. Hal tersebut terlihat dari adanya kata yang memiliki keterkaitan dengan lebih dari satu isu. Selain itu, data Instagram cenderung tidak terstruktur dan dapat mengandung kata tidak baku, singkatan, variasi penulisan, serta istilah khusus yang dapat memengaruhi pembentukan topik. Oleh karena itu, kualitas *preprocessing*, termasuk *stemming* dan *stopword removal*, menjadi penting dalam menghasilkan representasi teks yang sesuai untuk pemodelan. LDA juga tidak secara langsung mempertimbangkan informasi interaksi seperti jumlah *likes* dan *comments* dalam pembentukan topik. Dengan demikian, hasil LDA dalam *x posting* pada setiap topik dan visualisasi *wordcloud* memberikan cara yang lebih mudah untuk melihat karakteristik hasil pemodelan. Namun, visualisasi tersebut tidak serta-merta menunjukkan bahwa topik dengan jumlah posting terbesar merupakan isu yang paling penting. Frekuensi kemunculan hanya menggambarkan seberapa banyak suatu isu dibahas dalam dataset yang digunakan. Penentuan prioritas atau tingkat urgensi isu memerlukan kriteria tambahan di luar hasil *topic modeling*. Dengan demikian, platform yang dikembangkan pada penelitian ini terutama berperan dalam mengumpulkan, mengelompokkan, dan memvisualisasikan kecenderungan isu publik, yang selanjutnya dapat menjadi dasar untuk analisis yang lebih mendalam.

Secara keseluruhan, hasil penelitian menunjukkan bahwa penggunaan LDA dengan lima topik mampu menghasilkan kelompok isu publik yang relatif dapat diinterpretasikan pada dataset Instagram, dengan nilai *C\_v coherence* sebesar

0,5237. Integrasi proses *crawling*, *preprocessing*, *topic modeling*, evaluasi, dan visualisasi dalam satu platform menjadi kontribusi utama penelitian karena memungkinkan proses identifikasi isu dilakukan secara lebih terintegrasi. Meskipun demikian, interpretasi hasil tetap perlu mempertimbangkan sifat probabilistik LDA, adanya *overlapping* antartopik, serta karakteristik data Instagram yang tidak terstruktur.

## 5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengembangkan Platform Analisis Isu Publik pada Media Sosial Instagram Menggunakan Pendekatan Topic Modeling dan Text Mining berbasis web menggunakan framework Streamlit. Platform yang dikembangkan mampu mengintegrasikan seluruh tahapan analisis, mulai dari *crawling* data Instagram, manajemen dataset, *preprocessing* teks, analisis topik menggunakan metode Latent Dirichlet Allocation (LDA), hingga visualisasi hasil analisis dalam satu lingkungan kerja yang terintegrasi. Hasil implementasi menunjukkan bahwa seluruh modul dapat berfungsi sesuai dengan kebutuhan sistem sehingga proses analisis data media sosial menjadi lebih praktis, sistematis, dan efisien tanpa memerlukan penggunaan beberapa perangkat lunak secara terpisah. Platform yang dihasilkan diharapkan dapat mendukung kegiatan penelitian maupun analisis isu publik berbasis data serta menjadi dasar pengembangan sistem analisis media sosial yang lebih komprehensif.

Pengembangan platform pada penelitian selanjutnya dapat diarahkan pada penambahan dukungan terhadap berbagai sumber data media sosial selain Instagram, seperti X, Facebook, dan TikTok, sehingga cakupan analisis menjadi lebih luas. Selain itu, platform dapat dikembangkan dengan mengintegrasikan metode *topic modeling* yang lebih mutakhir, seperti BERTopic atau Top2Vec, serta menambahkan fitur visualisasi interaktif, analisis tren temporal, dan integrasi metode pengambilan keputusan untuk mendukung proses penentuan prioritas isu publik secara lebih komprehensif.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Direktorat Riset, Teknologi, dan Pengabdian kepada Masyarakat (DRTPM), Kementerian Pendidikan Tinggi, Sains, dan Teknologi Republik Indonesia atas dukungan pendanaan melalui skema Penelitian Dosen Pemula (PDP) sehingga penelitian ini dapat terlaksana dengan baik.

## DAFTAR PUSTAKA

- [1] G. S. Buana, R. Tyasnurita, N. C. Puspita, R. A. Vinarti, and F. Mahananto, "Text-Based Content Analysis on Social Media Using Topic Modelling to Support Digital Marketing,"

- JOIV : Int. J. Inform. Visualization*, vol. 8, no. 1, p. 88, Mar. 2024, doi: 10.62527/joiv.8.1.1636.
- [2] I. Z. Mustaqim and R. R. Suryono, "A Systematic Literature Review of Topic Modeling Techniques in User Reviews," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 11, no. 2, pp. 238–253, Jul. 2025, doi: 10.20473/jisebi.11.2.238-253.
- [3] A. M. G. A., S. Robledo, and M. Zuluaga, "Topic Modeling: Perspectives From a Literature Review," *IEEE Access*, vol. 11, pp. 4066–4078, 2023, doi: 10.1109/ACCESS.2022.3232939.
- [4] Cici Suhaeni, L. N. A. Mualifah, and H. Wijayanto, "LDA Topic Modeling Analysis of Public Discourse on Indonesia's Free Nutritious Meals Program (MBG)," *IJID*, vol. 14, no. 1, pp. 587–600, Jun. 2025, doi: 10.14421/ijid.2025.5211.
- [5] D. Choirinnisa *et al.*, "LDA Topic Modeling: Twitter-Based Public Opinion on Indonesian Ministry of Finance," *Sinkron*, vol. 9, no. 2, pp. 849–863, May 2025, doi: 10.33395/sinkron.v9i2.14719.
- [6] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research* 3, Jan. 2003.
- [7] M. Röder, A. Both, and A. Hinneburg, "Exploring the Space of Topic Coherence Measures," in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, Shanghai China: ACM, Feb. 2015, pp. 399–408. doi: 10.1145/2684822.2685324.
- [8] I. Vayansky and S. A. P. Kumar, "A review of Topic Modeling Methods," *Information Systems*, vol. 94, p. 101582, Dec. 2020, doi: 10.1016/j.is.2020.101582.
- [9] N. Hikmah, "Pemanfaatan Text Mining dalam Pencarian Ayat AlQuran menggunakan TF-IDF dan Cosine Similarity," *Jurnal Antartika*, vol. 08, no. 01, pp. 13–22, Apr. 2018.
- [10] R. Egger and J. Yu, "A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts," *Front. Sociol.*, vol. 7, p. 886498, May 2022, doi: 10.3389/fsoc.2022.886498.
- [11] Hastuti, H. F. Maulana, and E. Satria, "Sentiment Analysis and Topic Modeling of Twitter Conversations in Indonesia's 2024 Presidential Election," *Pekommas*, vol. 10, no. 1, Jun. 2025, doi: 10.56873/jpkm.v9i1.5545.
- [12] Y. Saputri, F. Syaki, and N. Hadinata, "Sentiment Analysis of Trending Topics on Social Media X Using Natural Language Processing and LSTM," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 6, pp. 3032–3041, Dec. 2025.
- [13] M. R. Fahlevvi and A. Sn, "Topic Modeling on Online News Portal Using Latent Dirichlet Allocation (LDA)," *Indonesian J. Comput. Cybern. Syst.*, vol. 16, no. 4, p. 335, Oct. 2022, doi: 10.22146/ijccs.74383.
- [14] D. Yu and B. Xiang, "Discovering topics and trends in the field of Artificial Intelligence: Using LDA topic modeling," *Expert Systems with Applications*, vol. 225, p. 120114, Sep. 2023, doi: 10.1016/j.eswa.2023.120114.