

BAB I

PENDAHULUAN

Bab ini membahas terkait permasalahan dan arah penelitian tugas akhir ini. Bab ini berisikan latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

1.1 Latar Belakang

Pada era digital saat ini, lanskap media telah bertransformasi secara fundamental. Media berita daring telah menjadi salah satu sumber utama bagi masyarakat untuk mengakses informasi dan memantau isu-isu terkini. Pergeseran konsumsi informasi ini menghadirkan tantangan sekaligus peluang bagi organisasi berita, terutama media lokal seperti Padang TV, agar tetap relevan dan kompetitif dalam menyajikan berita terkini. Media tidak hanya dituntut untuk menyajikan berita secara cepat, tetapi juga harus akurat dalam mengidentifikasi topik yang relevan dan menarik perhatian audiens (Rakhmawati et al., 2024). Tantangan terbesarnya adalah bagaimana memahami topik yang sedang hangat dibicarakan publik di tengah derasnya arus informasi.

Seiring dengan pergeseran fenomena tersebut, media sosial telah menjadi wadah utama untuk penyebaran informasi dan diskusi publik dalam membentuk opini (Basher et al., 2025). Platform media sosial seperti Instagram, YouTube, X, dan TikTok kini menjadi sumber data primer untuk memahami diskursus publik (Wahyuni et al., 2025). Di antara platform tersebut, Instagram telah berevolusi dari sekadar layanan berbagi foto menjadi saluran interaksi utama bagi banyak organisasi dan lembaga (Abidin et al., 2021).

Berbagai organisasi yang terlibat dalam komunikasi publik, termasuk pemerintah dan lembaga layanan sosial menggunakan Instagram untuk menyebarkan informasi, menerima tanggapan masyarakat, dan membangun percakapan dua arah melalui komentar, *likes*, *direct message*, serta unggahan ulang (Abidin et al., 2021). Hal tersebut menunjukkan bahwa Instagram tidak hanya berfungsi sebagai saluran distribusi konten, tetapi juga sebagai sumber data penting untuk menganalisis opini publik, tema percakapan, dan kecenderungan isu yang berkembang di masyarakat.

Interaksi yang terekam dalam bentuk komentar, jumlah *likes*, *caption*, dan waktu unggah pada platform media sosial menyediakan jejak digital yang kaya untuk dianalisis secara kuantitatif maupun kualitatif (Wahyuni et al., 2025). Data ini telah terbukti efektif digunakan sebagai sumber data utama untuk memetakan aspek opini publik dan topik yang sedang diperbincangkan (Basher et al., 2025). Namun, dalam praktiknya data ini belum dikelola dengan optimal.

Kebutuhan akan pengelolaan data ini sangat relevan bagi objek penelitian, yaitu Padang TV. Sebagai salah satu stasiun televisi lokal di Sumatera Barat, Padang TV mendedikasikan porsi siaran yang signifikan untuk program berita, yakni sebesar 56% dari total program yang didorong oleh tingginya kebutuhan masyarakat akan informasi terkini. proses sentral yang mengendalikan bisnis mereka adalah manajemen produksi siaran berita, khususnya pada tahap perencanaan atau pra-produksi. Pada tahap ini, tim redaksi menyelenggarakan rapat redaksi (*agenda setting*) untuk menentukan topik, narasumber, dan wilayah peliputan berita. Penentuan topik pada proses *agenda setting* diprioritaskan pada *headline* (berita umum) berkaitan dengan isu-isu publik baik secara nasional maupun daerah (Faudzi & Fajri, 2024). Isu-isu publik tersebut dapat diketahui melalui konten media sosial pada platform Instagram, di mana tren topik bahasan yang diekstrak dapat digunakan oleh tim redaksi sebagai acuan data yang objektif untuk mendukung pengambilan keputusan editorial. Oleh karena itu, optimalisasi proses *agenda setting* melalui dukungan pengelolaan data komputasional merupakan fokus utama dalam penelitian ini.

Meskipun proses *agenda setting* krusial, ketergantungan pada mekanisme rapat redaksi memiliki kelemahan yang mendasar. Berdasarkan temuan Faudzi & Fajri (2024), seluruh tahapan pra-produksi, mulai dari perencanaan hingga persiapan topik, dilakukan melalui diskusi saat rapat redaksi. Seluruh tim yang terlibat akan mendiskusikan hasil temuan dan analisis isu publik melalui sosial media yang telah dilakukan sebelumnya. Tanpa dukungan alat analisis data, proses ini bergantung pada analisis masing-masing jurnalis yang bersifat subjektif dan rentan terhadap bias. Bagian redaksi berisiko melewatkan *insight* dan tren tersembunyi yang hanya dapat ditemukan melalui analisis data skala besar. Hal ini menciptakan kesenjangan antara tujuan ideal Padang TV untuk mengutamakan isu

kekinian yang berkembang di masyarakat dan proses operasional mereka yang belum didukung oleh analisis data yang komprehensif (Faudzi & Fajri, 2024).

Kesenjangan tersebut tidak hanya berdampak pada proses identifikasi isu yang berkembang, tetapi juga berpengaruh terhadap berbagai keputusan strategis redaksi. Berdasarkan hasil wawancara dengan pihak Padang TV, informasi mengenai topik yang sedang menjadi perhatian publik digunakan sebagai landasan dalam menentukan berita yang diprioritaskan untuk diliput, menentukan urutan penayangan berita, serta mengidentifikasi momentum pemberitaan yang tepat agar informasi tetap relevan ketika diterima masyarakat. Pada proses identifikasi topik yang sedang berjalan, beresiko terlewatnya topik yang sedang berkembang atau keterlambatan dalam merespons momentum suatu peristiwa. Kondisi ini berpotensi menurunkan kualitas berita yang dihasilkan karena redaksi menjadi kurang cepat dan akurat dalam merespons isu yang sedang berkembang. Akibatnya, strategi penerbitan berita menjadi kurang efektif sehingga dapat berdampak pada penurunan jumlah penonton (*viewer*) yang menghambat peningkatan jumlah pengikut (*follower*) media sosial, serta mengurangi daya saing Padang TV dalam menyajikan informasi yang relevan dan tepat waktu kepada masyarakat.

Penerapan teknologi *natural language processing* (NLP) merupakan solusi strategis untuk menjembatani kesenjangan tersebut. Selain kemampuannya menganalisis data berskala besar secara objektif, NLP juga dapat diintegrasikan dengan algoritma klasifikasi seperti *Logistic Regression* yang terbukti efektif dalam mengklasifikasikan teks berlabel banyak berkat pendekatannya yang bersifat probabilistik dan efisien pada data berdimensi tinggi (Gouliev & Perez-tellez, 2024). Teknik utama dalam bidang NLP adalah *topic modeling* (pemodelan topik), yaitu metode yang digunakan untuk mengungkap pola topik tersembunyi dari koleksi dokumen teks yang masif secara otomatis (Nursyahrina et al., 2024). Seiring dengan kompleksitas data yang terus meningkat, metode *topic modeling* pun mengalami evolusi dan menghasilkan berbagai pendekatan dengan karakteristik serta keterbatasan spesifiknya masing-masing.

Salah satu pendekatan mutakhir yang kini menjadi standar baru dalam *topic modeling* adalah BERTopic. BERTopic merupakan teknik pemodelan topik berbasis *deep learning* yang mengintegrasikan model bahasa *transformer* dan

class-based TF-IDF (c-TF-IDF) untuk menghasilkan kluster topik yang padat dan mudah diinterpretasikan (Grootendorst, 2022). Kemampuan utama BERTopic terletak pada pemanfaatan *embedding* dokumen untuk menangkap konteks semantik secara mendalam yang kemudian dipadukan dengan teknik reduksi dimensi UMAP dan algoritma klustering HDBSCAN. Mekanisme ini memungkinkan model untuk tidak hanya mengelompokkan dokumen berdasarkan kata kunci yang muncul bersamaan, tetapi juga memahami makna tersirat di balik penggunaan kata dalam berbagai konteks kalimat (Basher et al., 2025).

Jika dibandingkan dengan metode terdahulu yang lebih konvensional seperti *Latent Dirichlet Allocation* (LDA), BERTopic menawarkan pendekatan yang lebih adaptif terhadap karakteristik data teks modern. Metode LDA, meskipun telah lama menjadi metode standar untuk analisis topik di media daring yang bekerja dengan asumsi probabilistik "*Bag-of-Words*" yang mengabaikan urutan kata dan konteks semantik. Keterbatasan ini membuat LDA cenderung kurang optimal saat menangani data teks pendek yang heterogen dan minim kata kunci, seperti *caption* media sosial karena ketergantungannya pada frekuensi kemunculan kata semata (Wahyuni et al., 2025).

BERTopic hadir menyempurnakan kekurangan tersebut dengan memanfaatkan *pretrained transformer embedding* yang mampu mempertahankan nuansa semantik dan relasi sintaktik antar kata, bahkan dalam dokumen yang singkat. Keunggulan ini membuat BERTopic mampu menghasilkan topik yang lebih presisi dan koheren tanpa memerlukan pra-pemrosesan teks yang ekstensif seperti *stemming* atau *lemmatization* yang seringkali menghilangkan informasi penting (Samosir et al., 2022). Selain itu, penggunaan HDBSCAN dalam *pipeline* BERTopic memungkinkan penentuan jumlah topik secara dinamis berdasarkan densitas data, mengatasi masalah subjektivitas penentuan jumlah topik manual (nilai K) yang menjadi kendala utama pada metode LDA (Egger et al., 2022).

Karakteristik dan keunggulan BERTopic ini menjadikannya metode yang sangat relevan dan cocok untuk diterapkan dalam penelitian ini, khususnya untuk menganalisis *metadata* video Instagram. Data *caption* Instagram memiliki variasi panjang teks yang ekstrim sehingga membutuhkan model dengan kemampuan pemahaman konteks yang kuat agar tidak terjadi *overlapping* topik. Implementasi

BERTopic memungkinkan ekstraksi topik yang tidak hanya mengandalkan kata kunci permukaan, tetapi juga menangkap esensi isu yang sedang dibahas sehingga dapat memberikan *insight* yang lebih akurat untuk mendukung proses *agenda setting* redaksi yang objektif.

Penelitian terdahulu telah menerapkan metode *topic modeling* BERTopic pada berbagai jenis data tekstual untuk melihat sejauh mana metode ini dapat digunakan. Sebagai contoh, Babalola et al. (2024) menggunakan BERTopic untuk memetakan topik pada teks pendek berupa *headline* berita berbahasa Inggris. Sementara itu, dalam konteks korpus bahasa Indonesia, Wahyuni et al. (2025) memanfaatkan metode ini untuk menganalisis opini masyarakat pada kolom komentar YouTube, dan Nursyahrina et al. (2024) menggunakannya untuk melacak tren publikasi ilmiah berdasarkan judul dan abstrak artikel.

Namun, pemanfaatan BERTopic dalam penelitian-penelitian tersebut masih terbatas pada analisis data media digital yang bersifat makro atau sekadar pemetaan tren dokumen secara umum. Sejauh ini, belum ditemukan penggunaan BERTopic yang diarahkan secara spesifik untuk menganalisis karakteristik *metadata* video Instagram dalam lingkup media penyiaran berita lokal. Celah penelitian inilah yang melandasi penerapan metode BERTopic dalam penelitian ini yang ditujukan untuk mengisi ruang kosong yang belum tersentuh oleh penelitian-penelitian terdahulu.

Meskipun BERTopic menawarkan kemampuan analisis yang lebih canggih, *output* dari metode ini perlu diterjemahkan ke dalam bentuk visual yang mudah dipahami oleh pengguna akhir agar hasil penerapannya optimal. Di sinilah peran *dashboard* analitik sebagai *user interface* (UI) menjadi krusial untuk menghubungkan sistem analitik kompleks dengan kebutuhan pengambil keputusan. Melalui *dashboard* tersebut, bagian redaksi tidak hanya dapat melakukan eksplorasi dinamis terhadap tren topik dan pola diskursus masyarakat secara objektif, tetapi juga dapat memberikan umpan balik langsung terhadap hasil sistem guna memastikan relevansi topik dengan konteks jurnalistik lokal. Penelitian Walingkas dan Saian (2023) menunjukkan bahwa *framework* Flask dapat diimplementasikan secara efektif untuk mengembangkan aplikasi *web* yang menyajikan data dan laporan analitik dalam format yang terstruktur dan interaktif.

Framework ini berbasis Python yang kompatibel dengan ekosistem NLP sehingga mempermudah integrasi model analitik seperti BERTopic ke dalam aplikasi *web* yang fungsional.

Urgensi penelitian ini muncul dari kesenjangan yang nyata antara potensi teknologi analisis data dan implementasi praktisnya di industri. Belum ada penelitian yang secara spesifik mengimplementasikan *topic modeling* berbasis *pipeline* BERTopic yang terbukti unggul dalam hal koherensi dan pemahaman semantik untuk mendukung pengambilan keputusan redaksional di media berita lokal seperti Padang TV. Oleh karena itu, penelitian ini bertujuan untuk menjawab kebutuhan tersebut melalui pembangunan sistem analisis topik yang mengintegrasikan BERTopic dengan *dashboard* interaktif berbasis Flask. Sistem ini dirancang untuk mengelola *metadata* konten video Instagram guna mengekstraksi topik relevan melalui *pipeline* BERTopic yang diperkuat dengan mekanisme verifikasi manusia (*human-in-the-loop*) untuk meningkatkan akurasi pelabelan secara adaptif menggunakan algoritma *Logistic Regression*. Dengan menyajikan visualisasi tren topik secara interaktif, sistem ini menjadi alat bantu pendukung keputusan bagi Padang TV dalam mendorong strategi penerbitan berita yang lebih objektif dan presisi berbasis data.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, masalah yang dirumuskan adalah bagaimana mengolah *metadata* konten video Instagram agar dapat memberikan informasi mengenai topik yang sedang berkembang secara lebih luas berbasis data, serta menyajikannya dalam sebuah sistem berbasis *web* yang mampu mendukung proses pengambilan keputusan redaksi dan strategi penerbitan berita di Padang TV.

1.3 Batasan Masalah

Berdasarkan rumusan masalah di atas, didapatkan batasan masalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah data tautan konten video berupa *feed* video dan *reels* Instagram yang diunggah oleh akun @info.minang, @info.padangpariaman, @info_pariaman, @infopadang_, @infosumbar, @kaba.bukittinggi, @infopasbar_, @pesisir.selatan,

@sudutpayakumbuh, @sudutlimapuluhkota, dan @salingkatanahdatar. Adapun *metadata* yang digunakan dalam penelitian ini adalah *metadata* yang didapatkan dari hasil ekstraksi data tautan konten video Instagram yang hanya meliputi *caption*, tanggal unggah, jumlah suka, dan jumlah komentar dalam rentang 13 Agustus 2025 hingga 12 Desember 2025 sehingga tidak mencakup komentar pengguna pada konten Instagram.

2. Metode pemodelan topik dalam penelitian ini dibatasi pada implementasi alur proses BERTopic menggunakan model *embedding* SBERT dengan bahasa Indonesia sebagai bahasa utama, UMAP, HDBSCAN, CountVectorizer, dan c-TF-IDF.
3. Penelitian ini mengadopsi alur kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM) dalam pengembangan model *natural language processing* (NLP) dan dibatasi hingga tahap *evaluation*.
4. Strategi pengembangan model menerapkan pendekatan *semi-supervised learning* dengan mengintegrasikan algoritma *Logistic Regression* sebagai model klasifikasi untuk melakukan pelabelan ulang pada *metadata* berdasarkan hasil verifikasi *human-in-the-loop* (HITL).
5. Struktur hierarki pengelompokan berita dalam sistem ini dibatasi pada dua tingkatan yang terdiri atas kategori sebagai pengelompokan utama dan topik sebagai bagian dari setiap kategori tersebut.
6. *Dataset* yang digunakan untuk melatih dan menguji model *Logistic Regression* dibatasi pada 362 *metadata* konten video Instagram yang telah melalui proses validasi oleh validator ahli guna menjaga kualitas label topik.
7. Sistem analisis topik berbasis *web* dibangun menggunakan *framework* Flask sebagai *backend* dan Bootstrap sebagai *frontend* dengan metode pengembangan sistem *Waterfall*. Ruang lingkup pengerjaan sistem dibatasi hingga tahap *construction* sehingga tidak mencakup tahap *deployment*.
8. Pengujian sistem analisis topik dilakukan menggunakan metode *black-box testing* dan *user acceptance testing* untuk memastikan kesesuaian sistem dengan kebutuhan Padang TV.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini sebagai berikut:

1. Menerapkan dan menganalisis *pipeline* pemodelan topik BERTopic untuk mengekstrak klaster-klaster topik dari *metadata* video Instagram.
2. Membangun mekanisme pelabelan ulang (*relabelling*) secara adaptif menggunakan algoritma *Logistic Regression* berbasis verifikasi *human-in-the-loop* (HITL) untuk meningkatkan akurasi pengelompokan topik.
3. Membangun sebuah sistem *dashboard* interaktif yang mampu memvisualisasikan topik yang ditemukan serta menganalisis tren topik dari waktu ke waktu.
4. Mengevaluasi model BERTopic dengan menggunakan *coherence score* dan *topic diversity* guna memvalidasi kualitas, koherensi, dan relevansi topik yang diekstraksi.
5. Mengevaluasi model *Logistic Regression* yang bertugas dalam klasifikasi menggunakan *confusion matrix*, *Brier score*, dan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat menghasilkan sebuah sistem analisis topik yang bermanfaat bagi Padang TV sebagai alat bantu pendukung keputusan berbasis data. Sistem yang dikembangkan mampu mengolah *metadata* video Instagram menjadi informasi mengenai topik yang sedang berkembang secara lebih luas dan terstruktur sehingga membantu pihak redaksi dalam mengidentifikasi tren konten publik yang relevan. Selain itu, sistem dapat menyajikan hasil analisis dalam bentuk visualisasi yang memudahkan pemantauan perkembangan topik serta mendukung penyusunan *agenda setting* dan strategi penerbitan berita sesuai dengan dinamika informasi terkini.

1.6 Sistematika Penulisan

Adapun sistematika penulisan laporan tugas akhir ini adalah sebagai berikut:

BAB I: PENDAHULUAN

Bab I memuat deskripsi mengenai konteks penelitian melalui latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan laporan.

BAB II: TINJAUAN PUSTAKA

Bab II menyajikan kajian literatur yang mencakup landasan teori terkait metode *machine learning* dan metode pengembangan sistem yang digunakan dalam penelitian.

BAB III: METODOLOGI PENELITIAN

Bab III menguraikan metodologi penelitian secara komprehensif, meliputi penentuan objek penelitian, metode pengumpulan data, dan tahapan-tahapan penelitian yang dilakukan.

BAB IV: ANALISIS DAN PERANCANGAN

Bab IV memfokuskan pada analisis dan perancangan sistem analisis topik. Bagian ini merinci setiap tahap penelitian, dimulai dari diagram perancangan sistem, perancangan *database*, dan perancangan *user interface* sistem.

BAB V: IMPLEMENTASI DAN HASIL

Bab V membahas implementasi sistem yang meliputi pengkodean program berdasarkan rancangan sistem dan pengujian sistem.

BAB VI: PENUTUP

Bab VI menyajikan penutup yang memuat simpulan hasil penelitian dan saran konstruktif untuk pengembangan sistem lebih lanjut di masa mendatang.