

Analisis Klasifikasi Teks Konten Instagram Ditsamapta Polda Kalimantan Tengah Menggunakan IndoBERT

Analysis of Text Classification for Instagram Content of Ditsamapta Central Kalimantan Regional Police Using IndoBERT

Maura Widyaningsih, Siti Maryamah, Gusniear Tulus Glorian

Program Studi Teknik Informatika, STMIK Palangkaraya, Jl. G. Obos No. 114, Palangka Raya, Kalimantan Tengah

maurawidya@gmail.com, sitiujian2018@gmail.com, tulusglorianz@gmail.com

ABSTRAK

Instagram menjadi salah satu kanal strategis yang digunakan oleh instansi pemerintah, termasuk Direktorat Samapta (Ditsamapta) Polda Kalimantan Tengah, untuk membangun komunikasi publik yang efektif dan humanis. Namun, evaluasi sistematis berbasis data terhadap kategori konten dan pola interaksi publik masih terbatas. Penelitian ini bertujuan untuk mengelompokkan teks caption dan komentar publik pada akun Instagram resmi Ditsamapta Polda Kalteng ke dalam kategori kelas data menggunakan metode IndoBERT. Data penelitian terdiri atas 50 unggahan dan 1.555 komentar publik yang diperoleh melalui proses web scraping dan prapemrosesan otomatis menggunakan Google Colab.

Model IndoBERT diterapkan untuk mengidentifikasi tiga kategori utama, yaitu operasional, patroli, dan lainnya. Hasil klasifikasi menunjukkan bahwa konten bertema operasional mendominasi unggahan ($\pm 58\%$), sedangkan kategori lainnya memiliki rata-rata engagement tertinggi, menandakan preferensi publik terhadap konten yang bersifat humanis dan sosial. Komentar publik mayoritas bersifat apresiatif, memperkuat citra positif dan efektivitas komunikasi kelembagaan Ditsamapta di media sosial.

Penelitian ini bersifat eksploratif dan berhasil membangun pipeline analisis berbasis *Natural Language Processing* (NLP) yang efisien dan replikabel. Hasilnya diharapkan menjadi dasar pengembangan strategi komunikasi publik berbasis data bagi institusi kepolisian serta memperluas penerapan teknologi *machine learning* dalam bidang komunikasi digital kelembagaan.

Kata kunci: *Instagram, IndoBERT, Klasifikasi Teks, Komunikasi Publik, NLP* .

ABSTRACT

Instagram has become a strategic communication channel for government institutions, including the Directorate of Samapta (Ditsamapta) of the Central Kalimantan Regional Police, in building effective and human-centered public communication. However, systematic data-driven evaluations of content categories and audience engagement patterns remain limited. This study aims to classify captions and public comments from the official Instagram account of Ditsamapta Polda Kalteng into thematic data categories using the IndoBERT method. The dataset consists of 50 posts and 1,555 public comments, collected through web scraping and automated preprocessing in Google Colab.

The IndoBERT model identified three main categories—operational, patrol, and others. The results show that operational content dominates the posts ($\pm 58\%$), while the “others” category achieved the highest average engagement, indicating a public preference for human-interest and social-oriented messages. Public comments were predominantly appreciative, reinforcing Ditsamapta’s institutional image and communication effectiveness on social media.

This exploratory study successfully developed an NLP-based analytical pipeline that is efficient and replicable. The findings are expected to provide a foundation for data-driven public communication strategies within law enforcement institutions and contribute to the broader application of machine learning in institutional digital communication research.

Keywords: *Instagram, IndoBERT, Text Classification, Public Communication, NLP*.

Pendahuluan

Media sosial telah menjadi salah satu sarana utama dalam membangun komunikasi publik, termasuk bagi lembaga pemerintahan seperti Kepolisian Negara Republik Indonesia (Ginting dkk., 2021). Instagram, sebagai salah satu platform visual terbesar, menawarkan kombinasi antara teks dan gambar yang memungkinkan penyampaian pesan publik secara humanis, kreatif, dan interaktif melalui berbagai format unggahan seperti foto, video, carousel, dan reels (Christoper, 2025).

Direktorat Samapta (Ditsampta) Polda Kalimantan Tengah merupakan salah satu satuan kerja yang secara aktif memanfaatkan Instagram sebagai media publikasi kegiatan operasional, himbauan keamanan, serta edukasi masyarakat. Aktivitas komunikasi digital ini berperan penting dalam membangun citra positif kepolisian dan meningkatkan kepercayaan publik. Namun demikian, pengelolaan konten yang dilakukan selama ini masih bersifat manual dan deskriptif, terbatas pada pengamatan terhadap jumlah tanda suka (*likes*) dan komentar (*comments*) (Yudananto dkk., 2023). Pendekatan ini belum mampu memberikan pemahaman mendalam mengenai pola komunikasi maupun karakteristik linguistik yang terbentuk dalam interaksi digital tersebut (Uricska, 2023).

Di sisi lain, teks yang terkandung dalam *caption* maupun komentar publik pada unggahan Instagram menyimpan informasi linguistik yang kaya dan dapat diolah secara komputasional. Melalui pendekatan *Natural Language Processing* (NLP), data teks dapat diproses untuk mengenali pola bahasa, tema pesan, serta hubungan semantik yang menggambarkan kecenderungan komunikasi publik. Salah satu model NLP yang memiliki performa unggul dalam konteks bahasa Indonesia adalah *IndoBERT* merupakan turunan dari arsitektur *Bidirectional Encoder Representations from Transformers* (*BERT*) yang mampu memahami konteks kata dalam dua arah (Amanda dkk., 2024). Kemampuan ini menjadikan *IndoBERT* efektif untuk menganalisis teks media sosial yang bersifat informal dan ringkas (Ahmadian dkk., 2024).

Dengan memanfaatkan *IndoBERT*, penelitian ini berupaya mengklasifikasikan teks *caption* dan komentar dari akun resmi Instagram Ditsampta Polda Kalimantan Tengah ke dalam kategori tematik tertentu. Pendekatan ini diharapkan dapat membantu lembaga dalam memahami pola komunikasi digitalnya secara lebih sistematis dan berbasis data. Selain itu, hasil penelitian ini diharapkan menjadi dasar bagi pengembangan strategi komunikasi publik kelembagaan yang lebih adaptif terhadap karakteristik audiens media sosial.

Berdasarkan latar belakang tersebut, penelitian ini berfokus pada penerapan *IndoBERT* untuk klasifikasi teks

konten media sosial kelembagaan. Metode ini diterapkan untuk mengklasifikasikan teks *caption* dan komentar pada akun Instagram Ditsampta, apa saja kategori kelas data yang dihasilkan dari proses klasifikasi tersebut. Hasil klasifikasi kategori tersebut dapat menggambarkan pola komunikasi digital Ditsampta di media sosial.

Berbagai penelitian sebelumnya menunjukkan efektivitas model *BERT* dan turunannya dalam mengolah teks media sosial berbahasa Indonesia. Murfi dkk. (2022) menunjukkan bahwa kombinasi *BERT* dengan arsitektur CNN dan RNN mampu meningkatkan performa klasifikasi teks dibandingkan metode tradisional. Saragih dan Manurung (2024) menegaskan konsistensi keunggulan *BERT* dalam berbagai konteks sosial media, sementara Shayamta dkk. (2025) membuktikan bahwa *IndoBERT* memberikan hasil yang seimbang bahkan lebih baik daripada *BiLSTM* pada teks pendek dan *noisy*. Temuan serupa dilaporkan oleh Khairani dkk. (2024) yang menunjukkan bahwa *IndoBERT* tetap akurat meski tanpa *stopword removal* atau *stemming*. Berdasarkan hasil-hasil tersebut, *IndoBERT* dipandang sebagai pendekatan yang relevan untuk analisis komunikasi digital publik melalui media sosial.

Tahapan penelitian yang dilakukan identifikasi masalah, pengumpulan data, pra-pemrosesan, klasifikasi teks dengan *IndoBERT*, analisis deskriptif dan interpretasi hasil. Fokus pada analisis klasifikasi teks media sosial kelembagaan, dengan data dikumpulkan secara melalui *web scraping* menggunakan pustaka *Instaloader* di lingkungan Google Colab. Pra-pemrosesan data mencakup pembersihan karakter non-alfabet, normalisasi teks, serta penyusunan data dalam format terstruktur agar siap dianalisis klasifikasi teks dengan model *IndoBERT*. Proses klasifikasi untuk mengelompokkan konten ke dalam tiga kategori utama, yaitu *operasional*, *patroli*, dan *lainnya*. Analisis deskriptif dan interpretasi hasil untuk meninjau distribusi kategori, pola interaksi publik, serta implikasinya terhadap strategi komunikasi digital lembaga.

Tujuan utama penelitian ini adalah mengelompokkan teks *caption* dan komentar publik ke dalam kategori kelas data menggunakan *IndoBERT*, serta menganalisis distribusi tema dan tingkat keterlibatan audiens. Secara praktis, penelitian ini memberikan manfaat berupa kerangka analisis berbasis data yang dapat membantu instansi publik memahami pola interaksi digital, sedangkan secara akademik, penelitian ini memperkuat kajian penerapan NLP dalam konteks komunikasi publik kelembagaan di Indonesia.

Metode

Penelitian ini menggunakan metode deskriptif kuantitatif dengan pendekatan komputasional berbasis NLP. Teori

NLP dengan model *Bidirectional Encoder Representations from Transformers (BERT)* sebagai pendekatan utama dalam analisis dan klasifikasi teks berbahasa Indonesia. Dalam konteks komunikasi digital, analisis teks digunakan untuk mengenali pola bahasa, struktur pesan, serta makna semantik yang terkandung dalam konten media sosial, khususnya pada *caption* dan *komentar (comment)* di platform Instagram.

Model *BERT* memiliki kemampuan untuk memahami konteks dua arah (*bidirectional context*) melalui mekanisme *self-attention*, yang memungkinkan sistem mengenali makna kata berdasarkan posisi dan hubungan antarkata dalam kalimat (Murfi dkk., 2022; Talaat, 2023). Versi adaptasi untuk bahasa Indonesia, *IndoBERT*, telah terbukti efektif dalam menangani teks pendek dan informal yang lazim digunakan di media sosial (Shayamta dkk., 2025).

1. Lokasi dan Sumber Data

Penelitian dilakukan pada akun resmi Instagram Direktorat Samapta Kepolisian Daerah Kalimantan Tengah (@ditsamapta_polda_kalteng). Akun ini dipilih karena merupakan kanal utama dalam penyebaran informasi publik, dokumentasi kegiatan operasional, dan imbauan keamanan kepada masyarakat (Perea dkk., 2021). Sumber data penelitian terdiri atas dua jenis:

- Data unggahan (*posts*), yang mencakup teks *caption*, waktu unggah (*timestamp*), jenis media (*foto*, *video*, *carousel*, atau *reels*), daftar *hashtag*, serta jumlah *likes* dan *komentar*.
- Data *komentar (comments)*, yang berisi teks *komentar publik* beserta waktu publikasi dan identitas pengguna yang terhubung dengan unggahan terkait.

Seluruh data diambil hanya dari konten yang bersifat publik sesuai dengan ketentuan *Terms of Use* dan *Privacy Policy* Instagram. Pengumpulan dilakukan menggunakan autentikasi akun resmi melalui *session file* agar proses lebih stabil dan terhindar dari pembatasan (*rate-limit*) *platform*.

Hasil scraping kemudian disimpan secara otomatis dalam format *Comma-Separated Values (CSV)* di direktori *Google Drive* penelitian. Penyimpanan dalam format terstruktur ini memudahkan tahap *pra-pemrosesan*, analisis dengan model *IndoBERT*, serta replikasi hasil pada penelitian berikutnya.

2. Tahapan Penelitian

Penelitian ini dilaksanakan melalui enam tahapan yang dilakukan secara sistematis dan terintegrasi sebagaimana

ditunjukkan pada Gambar 1. Setiap tahapan saling berhubungan membentuk *pipeline* analisis berbasis NLP yang digunakan untuk mengklasifikasikan teks konten Instagram Ditsamapta Polda Kalimantan Tengah.



Gambar 1 Tahapan Penelitian

a. Identifikasi Masalah

Tahapan awal dilakukan dengan mengidentifikasi isu utama dalam komunikasi publik kelembagaan, yaitu belum adanya evaluasi berbasis data terhadap pola bahasa dan karakter konten pada media sosial resmi Ditsamapta Polda Kalimantan Tengah. Berdasarkan hasil identifikasi tersebut, dirumuskan fokus penelitian untuk mengembangkan sistem klasifikasi teks menggunakan model *IndoBERT* sebagai alat bantu analisis komunikasi digital.

b. Pengumpulan Data

Data penelitian diperoleh melalui proses *web scraping* otomatis menggunakan pustaka *Instaloader* di lingkungan *Google Colab*. Data yang dikumpulkan mencakup teks *caption*, *komentar publik*, waktu unggahan (*timestamp*), jenis media (*foto*, *video*, *carousel*), serta jumlah *likes* dan *comments*. Semua data diambil dari konten publik sesuai kebijakan *Terms of Use* dan *Privacy Policy* Instagram.

c. Pra-pemrosesan Data

Data hasil *scraping* kemudian melalui proses *pra-pemrosesan* agar siap dianalisis oleh model NLP. Proses ini meliputi pembersihan karakter *non-alfabet*, penghapusan simbol dan emoji, konversi huruf menjadi *lowercase*, serta normalisasi format teks. Selanjutnya, data disusun ke dalam format CSV terstruktur yang memuat variabel penting seperti *post_id*, *caption_clean*, *hashtags_list*, dan *engagement_total*.

d. Klasifikasi Teks Menggunakan IndoBERT

Tahap utama analisis dilakukan dengan menerapkan model IndoBERT untuk mengelompokkan teks ke dalam tiga kategori tematik: operasional, patroli, dan lainnya. Proses klasifikasi dilakukan melalui tokenisasi, pembentukan representasi vektor, dan pemetaan konteks semantik dari teks *caption*. Hasil klasifikasi menggambarkan pola linguistik dan tema dominan dalam komunikasi digital Ditsamapta.

e. Analisis Deskriptif

Setelah klasifikasi selesai, dilakukan analisis deskriptif terhadap hasil keluaran model untuk menilai distribusi kategori konten, jumlah unggahan per kelas, serta hubungan antara kategori dengan tingkat keterlibatan (*engagement*). Tahapan ini juga mencakup deskripsi pola komentar publik untuk memperkaya pemahaman terhadap interaksi sosial yang terjadi.

f. Interpretasi Hasil

Tahap akhir penelitian adalah interpretasi hasil analisis untuk menarik kesimpulan terkait efektivitas komunikasi publik yang dilakukan Ditsamapta melalui Instagram. Hasil interpretasi digunakan untuk merumuskan rekomendasi strategi komunikasi berbasis data serta peluang pengembangan sistem evaluasi konten yang lebih otomatis dan berkelanjutan.

Seluruh proses di atas diimplementasikan secara terintegrasi dalam *pipeline* analisis di *Google Colab*, sehingga dapat direplikasi dan dikembangkan untuk penelitian sejenis di masa mendatang.

3. Teknik Pengumpulan Data

Proses pengumpulan data dilakukan menggunakan metode *web scraping* secara otomatis dengan bantuan pustaka *Instaloader*. Teknik ini memungkinkan peneliti mengakses metadata unggahan dan komentar publik dari akun Instagram tanpa mengganggu sistem internal platform. Pengambilan data dilakukan dalam periode 1–15 Juli 2025. Untuk menjaga stabilitas proses scraping dan menghindari *rate-limit* dari Instagram, peneliti menggunakan autentikasi akun resmi melalui *session file*.

Seluruh hasil pengumpulan data disimpan dalam format *Comma-Separated Values (CSV)* di direktori *Google Drive*, sehingga memudahkan proses prapemrosesan dan analisis selanjutnya menggunakan *IndoBERT*.

4. Teknik Analisis Data

Analisis data dilakukan dengan memanfaatkan model *IndoBERT* untuk melakukan klasifikasi teks *caption* dan interpretasi terhadap komentar publik (Muftie dan Haris, 2023), (Scudetto dan Romadhony, 2023) (Novandian dkk., 2024). Tahapan analisis meliputi:

- a. Pembersihan dan tokenisasi data teks, agar setiap kata dapat diidentifikasi secara kontekstual oleh model
- b. Penerapan *IndoBERT* untuk klasifikasi konteks, guna mengelompokkan teks ke dalam tiga kategori utama: *operasional*, *patroli*, dan *lainnya*.
- c. Analisis deskriptif hasil klasifikasi, dengan menampilkan distribusi kelas, proporsi setiap kategori, serta hubungan antara tema konten dan tingkat *engagement*.
- d. Interpretasi komentar publik, untuk melihat bagaimana respons masyarakat memperkuat atau menanggapi pesan yang disampaikan oleh Ditsamapta.

Hasil analisis disajikan dalam bentuk tabel, grafik, dan visualisasi deskriptif yang menggambarkan pola komunikasi digital, distribusi konten, dan efektivitas pesan publik melalui media sosial kepolisian.

Hasil Dan Pembahasan

I. Hasil Penelitian

Penelitian ini menghasilkan sistem klasifikasi teks berbasis IndoBERT untuk menganalisis konten Instagram Direktorat Samapta (Ditsamapta) Polda Kalimantan Tengah. Proses penelitian mencakup tiga tahap utama, yaitu akuisisi data, prapemrosesan teks, dan klasifikasi menggunakan model IndoBERT.

Tahap awal penelitian dilakukan melalui proses *web scraping* dengan pustaka *Instaloader* yang dijalankan di lingkungan *Google Colab*. Periode pengumpulan data ditetapkan pada tanggal 1–15 Juli 2025. Hasil akuisisi metadata berhasil menghimpun 50 unggahan dan 1.555 komentar publik, yang selanjutnya diekspor dalam format CSV sebagai dataset utama penelitian.

Gambar 2 merupakan contoh sampel metadata hasil dari akuisisi komentar.

ID	Username	Timestamp	Content	Engagement	Category
1	@ditamapta	2025-07-15 10:00:00	Ini adalah postingan pertama	100 likes, 5 comments	Operasional
2	@ditamapta	2025-07-15 10:00:00	Postingan kedua	150 likes, 10 comments	Patroli
3	@ditamapta	2025-07-15 10:00:00	Postingan ketiga	200 likes, 15 comments	Operasional
4	@ditamapta	2025-07-15 10:00:00	Postingan keempat	180 likes, 12 comments	Patroli
5	@ditamapta	2025-07-15 10:00:00	Postingan kelima	220 likes, 18 comments	Operasional
6	@ditamapta	2025-07-15 10:00:00	Postingan keenam	190 likes, 14 comments	Patroli
7	@ditamapta	2025-07-15 10:00:00	Postingan ketujuh	210 likes, 16 comments	Operasional
8	@ditamapta	2025-07-15 10:00:00	Postingan kedelapan	230 likes, 19 comments	Patroli
9	@ditamapta	2025-07-15 10:00:00	Postingan kesembilan	250 likes, 21 comments	Operasional
10	@ditamapta	2025-07-15 10:00:00	Postingan kesepuluh	270 likes, 23 comments	Patroli

Gambar 2. Metadata hasil akuisisi komentar

Gambar 2 menunjukkan contoh data komentar publik yang berisi emoji apresiatif, seperti dan , yang menandakan ekspresi dukungan terhadap konten unggahan.

Gambar 3 merupakan metadata hasil akuisisi postingan.



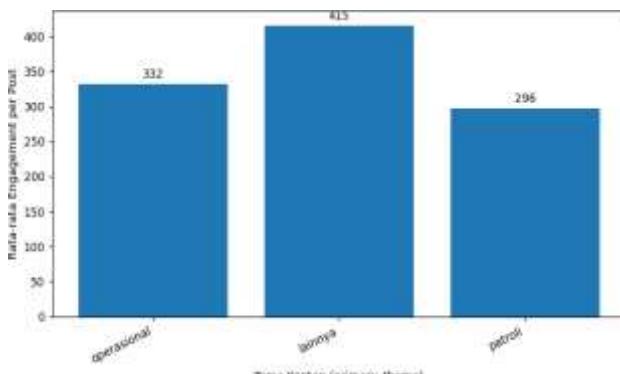
Gambar 3. Metadata Hasil Akuisisi Postingan

Gambar 3 menunjukkan metadata unggahan yang mencakup teks caption, jenis media, jumlah likes dan komentar, serta tautan unggahan.

Data unggahan diformat ulang untuk menghasilkan variabel penting seperti *post_id*, *timestamp*, *caption_clean*, *hashtags_list*, dan *engagement_total*. Proses prapemrosesan teks meliputi pembersihan karakter non-alfabet, penghapusan simbol, serta normalisasi huruf kecil agar teks siap diproses oleh model *IndoBERT*.

Selain analisis terhadap teks *caption*, komentar publik juga dianalisis secara deskriptif untuk memahami konteks interaksi pada setiap kategori konten. Komentar pada kategori operasional umumnya bersifat informatif dan singkat, menandakan bentuk perhatian publik terhadap kegiatan rutin kepolisian. Sebaliknya, komentar pada kategori lainnya lebih sering berisi ekspresi apresiatif berupa kata pujian, emoji, dan dukungan moral. Pola ini menunjukkan adanya keterlibatan emosional (*emotional engagement*) audiens yang lebih tinggi terhadap konten dengan pendekatan humanis dan sosial.

Gambar 4 merupakan hasil dari pemasaran perbandingan rata-rata Engagement dengan *IndoBert* dalam menentukan kategori konten.



Gambar 4. Perbandingan Rata-rata Engagement Berdasarkan Kategori Konten

Gambar 4 menunjukkan hasil pemrosesan dengan *IndoBERT*, di mana kategori “lainnya” memperoleh rata-rata engagement tertinggi.

Hasil utama klasifikasi teks *caption* menggunakan *IndoBERT* menunjukkan bahwa:

- Kategori Operasional mendominasi unggahan, mencakup 58% dari total data.
- Kategori Patroli menempati 26%, mencerminkan dokumentasi kegiatan preventif dan patroli harian.
- Kategori Lainnya mencakup 16% unggahan, namun memiliki rata-rata engagement tertinggi, yakni 178 interaksi per unggahan.

Pipeline penelitian, mulai dari akuisisi metadata, prapemrosesan, hingga klasifikasi semantik, berhasil dijalankan secara utuh. Hasil ini membuktikan bahwa *IndoBERT* dapat digunakan secara efektif untuk mengidentifikasi dan mengelompokkan kategori konten berbasis teks Instagram secara sistematis, efisien, dan replikabel.

2. Pembahasan

Berdasarkan hasil klasifikasi, ditemukan bahwa komunikasi digital pada akun Instagram Ditsampta Polda Kalimantan Tengah masih berfokus pada kategori operasional. Dominasi tema operasional sejalan dengan fungsi utama Ditsampta sebagai pelaksana kegiatan preventif dan patroli rutin di lapangan. Meskipun demikian, dari perspektif komunikasi publik, variasi konten menjadi aspek penting untuk menjaga keberagaman pesan dan memperluas jangkauan audiens.

Menariknya, kategori “lainnya” menunjukkan tingkat engagement tertinggi, yang mengindikasikan adanya preferensi publik terhadap narasi humanis, edukatif, dan sosial. Temuan ini konsisten dengan berbagai studi komunikasi digital yang menegaskan bahwa konten dengan elemen *human interest* mampu meningkatkan keterlibatan emosional audiens dan memperkuat citra positif lembaga publik.

Dari sisi teknis, hasil penelitian menunjukkan bahwa *IndoBERT* mampu mengenali konteks linguistik teks media sosial secara konsisten, bahkan ketika struktur kalimat bersifat *informal*. Model ini efektif dalam mengelompokkan teks ke dalam kategori tematik tanpa memerlukan proses *feature engineering* manual. Namun demikian, reliabilitas model masih bergantung pada keseimbangan jumlah data per kategori dan kualitas tahap prapemrosesan. Oleh karena itu, penelitian ini bersifat eksploratif, dan hasilnya diharapkan menjadi dasar pengembangan penelitian lanjutan dengan dataset yang lebih besar dan model *fine-tuned* khusus untuk konteks komunikasi publik.

Selain itu, hubungan antara jenis konten dan keterlibatan audiens memberikan wawasan strategis bagi institusi kepolisian. Konten operasional memang esensial untuk menunjukkan *transparansi* dan profesionalitas, tetapi konten bertema sosial dan humanis terbukti lebih efektif dalam memperkuat citra positif dan kedekatan dengan masyarakat. Dengan demikian, penelitian ini tidak hanya menghasilkan model klasifikasi teks, tetapi juga memberikan implikasi praktis bagi strategi komunikasi digital kelembagaan.

Simpulan Dan Saran

1. Simpulan

Penelitian ini berhasil menerapkan model IndoBERT untuk mengklasifikasikan teks caption dan komentar publik dari akun Instagram resmi Ditsamapta Polda Kalimantan Tengah. Melalui proses *web scraping* dan prapemrosesan di *Google Colab*, diperoleh 50 unggahan dan 1.555 komentar yang kemudian dianalisis ke dalam tiga kategori utama, yaitu operasional, patroli, dan lainnya.

Hasil klasifikasi menunjukkan bahwa kategori operasional mendominasi unggahan ($\pm 58\%$), sedangkan kategori lainnya memiliki *rata-rata engagement* tertinggi, mencerminkan ketertarikan publik terhadap konten yang bersifat humanis dan sosial. Komentar publik umumnya bersifat apresiatif, memperkuat citra positif institusi di ruang digital.

Secara keseluruhan, penelitian ini menghasilkan pipeline analisis berbasis *Natural Language Processing* (NLP) yang efektif dan dapat direplikasi oleh instansi pemerintah untuk mendukung evaluasi komunikasi publik berbasis data pada media sosial.

2. Saran

Penelitian selanjutnya disarankan untuk memperluas periode pengumpulan data agar hasil yang diperoleh lebih representatif dan mencerminkan dinamika komunikasi digital dalam jangka waktu yang lebih panjang. Selain itu, proses pelabelan manual (*gold set*) perlu dilakukan untuk meningkatkan akurasi model *IndoBERT* melalui *fine-tuning* yang lebih kontekstual terhadap bahasa Indonesia informal di media sosial. Pengembangan hasil penelitian juga dapat diarahkan pada pembuatan dashboard analisis berbasis web yang memudahkan tim pengelola media sosial dalam melakukan evaluasi konten secara otomatis dan real-time. Selanjutnya, penerapan metode serupa di platform lain seperti TikTok dan Facebook juga direkomendasikan guna memperluas pemahaman tentang efektivitas komunikasi publik lintas kanal digital dan memperkuat strategi komunikasi kelembagaan berbasis data.

Pustaka Acuan

- Ahmadian, H., Abidin, T., Riza, H., dan Muchtar, K. (2024). Hybrid Models for Emotion Classification and Sentiment Analysis in Indonesian Language. *Appl. Comput. Intell. Soft Comput.*
- Amanda, C., Jaya, I., dan Arisandi, D. (2024). Identification of Sexual Harassment in Social Media Comments Using IndoBERT and Support Vector Machine. *2024 8th International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM)*, 42-45.
- Christoper, A., dan Rukmana, O. (2025). Police Digital Narrative in the Transformation Era: Netnographic Analysis of Instagram ADS Using NVIVO. *Eduvest - Jurnal of Universal Studies*. <https://doi.org/10.59188/eduvest.v5i7.51487>.
- Ginting, M., Saragih, Y., dan Rohani, L. (2021). Efektivitas Penggunaan Media Sosial Instagram dalam Meningkatkan Citra Kepolisian Polsek Delitua. *Communication dan Social Media*.
- Khairani, U. dkk. (2024). Pengaruh Tahapan Preprocessing terhadap Model IndoBERT pada Komentar Instagram Berita.
- Muftie, F., dan Haris, M. (2023). IndoBERT Based Data Augmentation for Indonesian Text Classification. *2023 International Conference on Information Technology Research and Innovation (ICITRI)*, 128-132.
- Murfi, H. (2022). BERT-based Combination of CNN and RNN for Indonesian Sentiment Analysis.
- Nasution, A. dan Nobelson, H. (2022). Pengaruh Media Sosial Instagram dalam Komunikasi Publik: Kasus Polri, *Jurnal Komunikasi Publik*, Medan.
- Novandian, Y., Luthfiarta, A., Assyifa, D., Setiawan, J., Cahyaningrum, L., Althoff, N., Rahayu, M., Nugraha, A., dan , R. (2024). IndoBERT-based Indonesian Cyberbullying Detection with Multi-stage Labeling. *2024 International Seminar on Application for Technology of Information and Communication (iSemantic)*, 515-521.
- Perea, D., Bonsón, E., dan Bednárová, M. (2021). Citizen reactions to municipalities' Instagram communication. *Gov. Inf. Q.*, 38, 101579.
- Saragih, H. dan Manurung, J. (2024). Leveraging the BERT Model for Enhanced Sentiment Analysis in Social Media.
- Scudetto, N., dan Romadhony, A. (2023). Image Caption Validation for Public Complaints on Social Media. *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)*, 1-6.
- Talaat, M. (2023). Sentiment Analysis Classification System Using Hybrid BERT Models.

- Uricska, E. (2023). Trust is in the (linguistic) details: policing digilect as a communication practice. *Információs Társadalom*.
- Yudananto, E., dan Damastuti, R. (2023). Analisis Isi Kuantitatif Akun Instagram @DivisihumasPolri dalam membangun Kepercayaan Publik. *POPULIKA*.