



## PROTEKSI ISI PROPOSAL

Dilarang menyalin, menyimpan, memperbanyak sebagian atau seluruh isi proposal ini dalam bentuk apapun kecuali oleh pengusul dan pengelola administrasi pengabdian kepada masyarakat

### PROPOSAL PENELITIAN 2025

Rencana Pelaksanaan Penelitian: 2025-09-01 s.d 2026-08-31

## 1. JUDUL PENELITIAN

Pengembangan Model Prediksi Keterlibatan Pengguna pada Postingan Instagram Fakultas Sains dan Teknologi Universitas Jambi Menggunakan Pembelajaran Mesin Berbasis Fitur Multimodal

Bidang Fokus Penelitian	Tema	Topik (Jika Ada)	Rumpun Bidang Ilmu
Teknologi Informasi dan Komunikasi	Aplikasi Teknologi Informasi dan Komunikasi	Aplikasi Teknologi Informasi dan Komunikasi	Kecerdasan Buatan

Skema Penelitian	Kategori SBK	Target Akhir TKT	Lama Penelitian (Tahun)
Penelitian Penugasan	SBK Riset Penugasan	1	1

## 2. IDENTITAS PENGUSUL

Nama, Peran	NIDN	Jabatan Fungsional	Perguruan Tinggi/Institusi	Program Studi/Bagian	Bidang Tugas	ID Sinta
Jefri Marzal ketua	0002066805	Lektor Kepala	Universitas Jambi	Informatika	Memimpin dan mengkoordinasi seluruh tahapan penelitian. Merancang metode prediksi keterlibatan menggunakan algoritma pembelajaran mesin. Mengintegrasikan hasil kerja tim, melakukan optimasi model final, menulis laporan penelitian lengkap, dan menyusun artikel ilmiah untuk publikasi jurnal nasional.	<a href="#">6005761</a> (0)

Muhammad Razi A. anggota	1030049004	Asisten Ahli	Universitas Jambi	Sistem Informasi	Mengumpulkan data keseluruhan postingan Instagram @fst_unja periode 2023-2025. Mengekstrak caption, gambar, timestamp, jumlah like, komentar, dan share. Memetakan setiap postingan dengan kalender akademik termasuk periode pendaftaran, ujian, wisuda, dan kegiatan fakultas.	<a href="#">6746109</a> (0)
Miranty Yudistira anggota	1994031702	PNS Non Fungsional	Universitas Jambi	Sistem Informasi	Membersihkan dan mengolah data tekstual caption. Menganalisis penggunaan bahasa Indonesia, hashtag populer, emoji, dan melakukan analisis sentimen. Mengidentifikasi pola kata kunci dan karakteristik teks yang mendapat respon positif dari followers fakultas.	<a href="#">6929978</a> (0)
Akhiyar Waladi anggota	1995072301	PNS Non Fungsional	Universitas Jambi	Teknik Elektro	Menganalisis karakteristik visual postingan meliputi komposisi gambar, warna dominan, pencahayaan, dan deteksi wajah. Mengidentifikasi pola waktu temporal posting optimal dan hubungan antara jenis konten visual dan pengaruhnya kepada keterlibatan audiens.	<a href="#">6931115</a> (0)
Hamzah Alghifari anggota	1998060404	CPNS Non Fungsional	Universitas Jambi	Sistem Informasi	Menguji akurasi model prediksi menggunakan berbagai metrik evaluasi. Membuat visualisasi hasil penelitian dalam bentuk grafik dan tabel. Mengembangkan aplikasi prototipe sederhana agar tim komunikasi fakultas dapat memprediksi kesuksesan postingan Instagram.	<a href="#">6965866</a> (0)

Ammar As'ad Nabilah Mahasiswa Bimbingan	F1E122011	-	Sains dan Teknologi	Sistem Informasi	Bertanggung jawab pada validasi dataset Instagram dan kalender akademik, memeriksa kelengkapan serta konsistensi data, lalu membuat catatan perubahan. Mendukung tim rekayasa fitur dengan menyusun dokumentasi teknis pengolahan data dan menyiapkan laporan parsial terstruktur.	-
INTAN SRI AYU BINTI DARUL NIZSHAM Mahasiswa Bimbingan	F1E124053	-	Sains dan Teknologi	Sistem Informasi	Mengembangkan antarmuka prototipe menggunakan Streamlit, mengintegrasikan API, serta melakukan uji penerimaan pengguna bersama Humas. Membuat panduan singkat penggunaan aplikasi, mendesain visualisasi hasil penelitian, dan mengelola arsip kode serta dataset secara sistematis untuk keberlanjutan penelitian.	-

### 3. MITRA KERJASAMA PENELITIAN (JIKA ADA)

Pelaksanaan penelitian dapat melibatkan mitra kerjasama yaitu mitra kerjasama dalam melaksanakan penelitian, mitra sebagai calon pengguna hasil penelitian, atau mitra investor

Mitra	Nama Mitra	Dana
Tidak Ada Mitra		

### 4. LUARAN DAN TARGET CAPAIAN

Luaran Wajib

Tahun Luaran	Jenis Luaran	Status target capaian	Keterangan
1	rumusan kebijakan yang monumental dalam bentuk arahan/kertas kebijakan (policy brief/policy paper)	draft	-
2	rumusan kebijakan yang monumental dalam bentuk arahan/kertas kebijakan (policy brief/policy paper)	publish	-
1	Artikel Ilmiah Pada Jurnal Nasional Terakreditasi SINTA 2	submit	-
2	Artikel Ilmiah Pada Jurnal Nasional Terakreditasi SINTA 2	submit	-
3	Artikel Ilmiah Pada Jurnal Nasional Terakreditasi SINTA 2	diterima	-
1	Prosiding Konferensi/Seminar Nasional Ber-ISBN/ISSN	diterima	-

## 5. ANGGARAN

Rencana Anggaran Biaya penelitian mengacu pada PMK dan buku Panduan Penelitian dan Pengabdian kepada Masyarakat yang berlaku.

Kelompok	Komponen	Item	Satuan	Harga Satuan	Volume	Total
<b>Bahan</b>						
	ATK	ATK (Kertas A4, tinta printer, alat tulis, map, dll)	Bulan	Rp. 150,000.00	12	Rp. 1,800,000.00
	Bahan Penelitian	USB/flashdisk 64GB, external HDD 2TB	Paket	Rp. 1,500,000.00	1	Rp. 1,500,000.00
	Bahan Persediaan	Software license dasar, domain website	Paket	Rp. 1,700,000.00	1	Rp. 1,700,000.00
<b>Total Anggaran Bahan</b>						<b>Rp. 5,000,000.00</b>
<b>Pengumpulan Data</b>						
	FGD Persiapan Penelitian	FGD Persiapan Penelitian dengan Tim Humas FST	Kali	Rp. 500,000.00	3	Rp. 1,500,000.00
	HR Pembantu Peneliti	HR Pembantu Peneliti (ekstraksi data manual)	OB	Rp. 500,000.00	4	Rp. 2,000,000.00
	HR Sekretaris/Administrasi Penelitian	HR Sekretaris/Administrasi Penelitian	OB	Rp. 300,000.00	6	Rp. 1,800,000.00
	Transport	Transport lokal (koordinasi tim)	Kali	Rp. 150,000.00	12	Rp. 1,800,000.00
	Uang harian rapat di dalam kantor	Uang Harian rapat di dalam kantor	OK	Rp. 50,000.00	10	Rp. 500,000.00
	Biaya Konsumsi	Biaya Konsumsi rapat koordinasi	Kali	Rp. 200,000.00	10	Rp. 2,000,000.00
	HR Pembantu Lapangan	HR Pembantu Lapangan (dokumentasi)	OB	Rp. 200,000.00	2	Rp. 400,000.00
<b>Total Anggaran Pengumpulan Data</b>						<b>Rp. 10,000,000.00</b>
<b>Sewa Peralatan</b>						
	Peralatan penelitian	Peralatan Penelitian (laptop khusus, kamera)	Bulan	Rp. 500,000.00	6	Rp. 3,000,000.00
	Obyek Penelitian	Obyek Penelitian (Instagram API credits)	Bulan	Rp. 500,000.00	12	Rp. 6,000,000.00
	Ruang penunjang penelitian	Ruang Penunjang Penelitian (cloud computing/GPU)	Bulan	Rp. 750,000.00	6	Rp. 4,500,000.00
	Transport Penelitian	Transport Penelitian (kendaraan operasional)	OK (kali)	Rp. 150,000.00	10	Rp. 1,500,000.00
<b>Total Anggaran Sewa Peralatan</b>						<b>Rp. 15,000,000.00</b>
<b>Analisis Data</b>						
	HR Sekretaris/Administrasi Penelitian	HR Sekretaris/Administrasi Penelitian	OB	Rp. 300,000.00	4	Rp. 1,200,000.00

Kelompok	Komponen	Item	Satuan	HargaSatuan	Volume	Total
	HR Pengolah Data	HR Pengolah Data (preprocessing & cleaning)	OP (Penelitian)	Rp. 600,000.00	6	Rp. 3,600,000.00
	Honorarium narasumber	Honorarium Narasumber (konsultasi ahli ML)	OJ	Rp. 750,000.00	3	Rp. 2,250,000.00
	Transport Lokal	Transport Lokal (konsultasi ahli)	OK (Kali)	Rp. 150,000.00	3	Rp. 450,000.00
	Biaya Konsumsi Rapat	Biaya Konsumsi Rapat analisis	OH	Rp. 100,000.00	5	Rp. 500,000.00
<b>Total Anggaran Analisis Data</b>						<b>Rp. 8,000,000.00</b>
<b>Pelaporan, Luaran Wajib dan Luaran Tambahan</b>						
	HR Sekretaris/Administrasi Penelitian	HR Sekretaris/Administrasi Penelitian	OB	Rp. 300,000.00	4	Rp. 1,200,000.00
	Uang harian rapat di dalam kantor	Uang Harian rapat di dalam kantor	OH	Rp. 50,000.00	10	Rp. 500,000.00
	Biaya Konsumsi Rapat	Biaya Konsumsi Rapat finalisasi	OH	Rp. 200,000.00	5	Rp. 1,000,000.00
	Biaya Seminar Nasional	Biaya Seminar Nasional (registrasi + akomodasi)	Paket	Rp. 750,000.00	2	Rp. 1,500,000.00
	Biaya Publikasi Artikel di Jurnal Nasional	Biaya Publikasi Artikel di Jurnal Nasional SINTA 2	Paket	Rp. 3,500,000.00	1	Rp. 3,500,000.00
	Biaya Luaran Publikasi Lainnya (HKI:Paten/Paten Sederhana, Hak Cipta dan yang sejenis yang dapat dikatakan sebagai HKI)	Biaya Luaran Publikasi Lainnya (prototipe aplikasi)	Paket	Rp. 3,000,000.00	1	Rp. 3,000,000.00
	Biaya Luaran Publikasi Lainnya (HKI:Paten/Paten Sederhana, Hak Cipta dan yang sejenis yang dapat dikatakan sebagai HKI)	Biaya Penerjemahan/Proofreading Artikel	Paket	Rp. 1,300,000.00	1	Rp. 1,300,000.00
<b>Total Anggaran Pelaporan, Luaran Wajib dan Luaran Tambahan</b>						<b>Rp. 12,000,000.00</b>

Kelompok	Komponen	Item	Satuan	Harga	Satuan Volume	Total
Total Anggaran						Rp. 50,000,000.00

6. Lokasi Penelitian

Laboratorium Komputasi Sains Fakultas Sains dan Teknologi Universitas Jambi, Kec. Jambi Luar Kota, Kab. Muaro Jambi, Prop. Jambi
--



## Isian Substansi Proposal

### SKEMA PENELITIAN PENUGASAN

Petunjuk: Pengusul hanya diperkenankan mengisi di tempat yang telah disediakan sesuai dengan petunjuk pengisian dan tidak diperkenankan melakukan modifikasi template atau penghapusan di setiap bagian.

#### JUDUL

Tuliskan Judul Usulan

Pengembangan Model Prediksi Keterlibatan Pengguna pada Postingan Instagram Fakultas Sains dan Teknologi Universitas Jambi Menggunakan Pembelajaran Mesin Berbasis Fitur Multimodal

#### RINGKASAN

Ringkasan penelitian tidak lebih dari 300 kata yang berisi urgensi, tujuan, dan luaran yang ditargetkan.

Media sosial Instagram telah menjadi platform komunikasi digital utama bagi institusi pendidikan tinggi dalam menjangkau mahasiswa dan masyarakat luas. Fakultas Sains dan Teknologi Universitas Jambi mengelola akun Instagram **@fst\_unja** sebagai **kanal komunikasi** untuk diseminasi informasi akademik, prestasi mahasiswa, dan kegiatan fakultas. Namun institusi menghadapi **tantangan signifikan** dalam mengoptimalkan strategi konten akibat **ketidakmampuan memprediksi tingkat keterlibatan audiens**. Analisis preliminer terhadap **227 postingan menunjukkan variabilitas ekstrem** dimana postingan dengan karakteristik serupa dapat menghasilkan perbedaan keterlibatan yang sangat signifikan, dengan mayoritas konten memiliki tingkat keterlibatan rendah sementara sebagian kecil mencapai viralitas tanpa pola yang dapat diprediksi sebelumnya.

Penelitian ini mengembangkan **model prediksi keterlibatan menggunakan pembelajaran mesin** yang mengintegrasikan **tiga dimensi fitur** utama secara komprehensif. Pertama, fitur tekstual mencakup analisis sentimen, kompleksitas linguistik, dan karakteristik hashtag berbahasa Indonesia. Kedua, fitur visual meliputi histogram warna, deteksi wajah, dan komposisi gambar. Ketiga, fitur temporal dengan pengkodean siklik untuk menangkap pola publikasi dan konteks kalender akademik. Metodologi menggunakan **algoritma XGBoost dan Random Forest** dalam arsitektur ensemble dengan dataset 227 postingan yang dibagi secara acak stratifikasi untuk pelatihan, validasi, dan pengujian.

Luaran penelitian meliputi: 1) model prediksi keterlibatan berbasis pembelajaran mesin ensemble untuk konten Instagram akademik; 2) dataset tervalidasi postingan Instagram institusi pendidikan dengan anotasi lengkap; 3) aplikasi prototipe prediksi keterlibatan yang mudah digunakan untuk operator media sosial; 4) publikasi pada jurnal nasional terakreditasi SINTA; dan 5) presentasi hasil penelitian pada seminar nasional. **Kebaruan penelitian** terletak pada **pengembangan model ensemble pertama yang mengintegrasikan fitur multimodal untuk konteks akademik Indonesia** dengan dataset terbatas, adaptasi pemrosesan bahasa alami untuk bahasa Indonesia akademik, dan optimasi model untuk implementasi praktis di institusi pendidikan. Penelitian mendukung pencapaian Indikator Kinerja Utama perguruan tinggi khususnya IKU 5 tentang diseminasi hasil karya dosen dan IKU 7 mengenai pembelajaran kolaboratif, sejalan dengan Rencana Strategis Universitas Jambi dalam transformasi digital.

#### KATA KUNCI

Kata kunci maksimal 5 kata

Prediksi Keterlibatan; Pembelajaran Mesin; Instagram; Komunikasi Digital; Fitur Multimodal

#### PENDAHULUAN

Penelitian Penugasan merupakan salah satu bentuk program pendanaan riset yang diarahkan untuk peningkatan produktifitas publikasi dan KI serta komersialisasi/ hilirisasi produk hasil riset. Program ini juga merupakan salah satu instrumen untuk mengakselerasi tercapainya kebijakan BRIN di sepuluh Bidang Fokus Riset yaitu: (1) Pangan-Pertanian, (2) Energi-energi Baru dan Terbarukan, (3) Kesehatan-

Obat, (4) Transportasi, (5) Teknologi Informasi dan Komunikasi, (6) Pertahanan dan Keamanan, (7) Material Maju, (8) Kemaritiman, (9) Kebencanaan, dan (10) Sosial Humaniora-Seni Budaya-Pendidikan. Program ini mengacu kepada RIP Universitas Jambi. Skema ini juga merupakan bentuk program pendanaan riset yang dibutuhkan UNJA untuk menghadapi issue-isue strategik, dimana hasil riset sangat dibutuhkan untuk peningkatan kinerja Universitas dan penyelesaian masalah strategik. Tujuan riset ini membentuk atau menguatkan kerjasama riset lintas disiplin antara perguruan tinggi dengan institusi riset atau industri pada suatu bidang dari 10 bidang fokus dan menghasilkan produk ilmu pengetahuan, teknologi, seni, dan budaya yang siap diterapkan.

Pendahuluan penelitian tidak lebih dari 1200 kata yang terdiri dari:

- Latar belakang dan rumusan permasalahan yang akan diteliti
- Pendekatan pemecahan masalah
- *State of the art* dan kebaruan
- Peta jalan (*road map*) penelitian 5 tahun kedepan (jika dalam bentuk konsorsium harus dilengkapi dengan roadmap penelitian konsorsium)
- Sitasi disusun dan ditulis berdasarkan sistem nomor sesuai dengan urutan pengutipan, mengikuti format Vancouver.

## Latar Belakang

Platform media sosial telah menjadi komponen integral dalam ekosistem komunikasi digital institusi pendidikan tinggi. Instagram sebagai platform visual dominan dengan lebih dari 2 miliar pengguna aktif bulanan global menawarkan peluang unik bagi institusi akademik untuk membangun keterlibatan dengan berbagai pemangku kepentingan (1). Di Indonesia, Instagram merupakan platform media sosial ketiga terpopuler setelah WhatsApp dan Facebook, dengan **penetrasi tinggi di kalangan usia 18-24 tahun** yang merupakan demografi utama mahasiswa (2).

Fakultas Sains dan Teknologi Universitas Jambi mengelola akun Instagram **@fst\_unja** sebagai kanal komunikasi digital utama untuk menjangkau mahasiswa, calon mahasiswa, alumni, dan masyarakat umum. Akun ini berperan penting dalam diseminasi informasi akademik, publikasi prestasi mahasiswa dan dosen, dokumentasi kegiatan fakultas, serta membangun citra institusi di ranah digital. Dengan total 227 postingan yang telah dipublikasikan, akun ini menjadi **representasi digital fakultas yang memerlukan strategi konten yang efektif** untuk memaksimalkan jangkauan dan keterlibatan audiens.

Tantangan utama yang dihadapi adalah **ketidakmampuan memprediksi tingkat keterlibatan** audiens pada konten Instagram. Observasi terhadap dataset postingan menunjukkan pola yang sangat tidak konsisten, dimana postingan dengan karakteristik serupa dapat menghasilkan tingkat keterlibatan yang berbeda secara dramatis. Sebagai contoh, postingan tentang prestasi mahasiswa terkadang viral dengan ribuan interaksi, namun di waktu lain hanya mendapat respons minimal meskipun kontennya serupa. Distribusi keterlibatan menunjukkan **ketimpangan yang signifikan** dimana sebagian kecil postingan mendominasi total keterlibatan sementara mayoritas konten tidak mendapat respons optimal. Fenomena ini menciptakan inefisiensi dalam alokasi sumber daya dimana tim komunikasi menghabiskan waktu dan usaha yang sama untuk setiap konten tanpa kepastian hasil.

Kompleksitas prediksi keterlibatan pada Instagram melibatkan interaksi berbagai faktor yang saling mempengaruhi. Penelitian menunjukkan bahwa karakteristik visual memiliki pengaruh dominan terhadap keterlibatan, dimana **keberadaan wajah manusia** dapat meningkatkan keterlibatan dan gambar berkualitas tinggi menghasilkan interaksi lebih banyak dibanding gambar berkualitas rendah (3). **Aspek tekstual** seperti panjang **caption**, penggunaan **emoji**, dan pemilihan **hashtag** juga berperan signifikan dalam **menentukan tingkat keterlibatan** postingan. Faktor **temporal** menunjukkan pola konsisten dimana **waktu publikasi** dapat mempengaruhi keterlibatan secara substansial, dengan waktu penerbitan unggahan yang optimal dapat **berbeda untuk setiap demografi** audiens.

Dalam konteks institusi pendidikan Indonesia, terdapat karakteristik unik yang membedakan dari pola umum media sosial komersial. **Kalender akademik** menciptakan **fluktuasi keterlibatan** dengan pola



musiman yang khas yaitu meningkat saat penerimaan mahasiswa baru dan wisuda, **menurun** saat **periode ujian**. **Preferensi konten** juga berbeda dimana audiens akademik lebih responsif terhadap **konten edukatif** dan prestasi dibanding hiburan semata. Penggunaan bahasa Indonesia dengan berbagai tingkat formalitas dan pencampuran istilah teknis menciptakan kompleksitas tersendiri dalam analisis tekstual yang memerlukan pendekatan khusus.

Gap penelitian yang teridentifikasi adalah **belum adanya model prediksi keterlibatan** yang spesifik untuk konteks institusi pendidikan Indonesia dengan dataset terbatas. **Penelitian** yang ada **umumnya fokus** pada **merek komersial** atau influencer dengan karakteristik **konten** dan **audiens** yang **berbeda** serta **dataset ribuan postingan**. Model yang dikembangkan untuk konteks berbahasa Inggris juga tidak dapat langsung diterapkan mengingat perbedaan struktur bahasa dan norma komunikasi. Penelitian ini dirancang untuk mengisi gap tersebut dengan mengembangkan model prediksi yang mempertimbangkan karakteristik unik konten akademik Indonesia dengan optimasi untuk dataset kecil.

**Pendekatan Pemecahan Masalah**

Penelitian mengadopsi pendekatan sistematis berbasis **pembelajaran mesin multimodal** yang dirancang untuk mengatasi kompleksitas prediksi keterlibatan Instagram dengan dataset terbatas namun representatif. Solusi dikembangkan dengan mempertimbangkan keterbatasan jumlah data yaitu 227 postingan dan kebutuhan praktis operator media sosial yang umumnya tidak memiliki latar belakang teknis mendalam. Pengumpulan data dilakukan secara komprehensif dari **akun resmi Instagram @fst\_unja** menggunakan kombinasi **Instagram Graph API** untuk metadata dan ekstraksi manual untuk metrik keterlibatan detail. Dataset 227 postingan mencakup periode tiga tahun terakhir yang menangkap variasi musiman dan evolusi pola keterlibatan. Data dibagi secara acak dengan pengambilan sampel berstrata untuk memastikan representasi yang proporsional yaitu 159 postingan untuk pelatihan, 34 postingan untuk validasi, dan 34 postingan untuk pengujian. Pembagian acak dipilih alih-alih pembagian temporal untuk menghindari bias waktu dan memastikan model dapat memprediksi berbagai kondisi temporal.

**Ekstraksi fitur** dilakukan pada tiga dimensi utama dengan total 55 fitur yang dirancang untuk menangkap aspek komprehensif dari setiap postingan. **Dimensi tekstual** menggunakan pemrosesan bahasa alami untuk bahasa Indonesia dengan library Sastrawi untuk stemming dan analisis sentimen yang diadaptasi untuk konteks lokal. **Dimensi visual** menggunakan teknik computer vision untuk ekstraksi histogram warna RGB/HSV, deteksi wajah dengan Haar Cascade, dan analisis komposisi gambar. **Dimensi temporal** mempertimbangkan tidak hanya waktu posting tetapi juga konteks kalender akademik dan acara fakultas dengan pengkodean siklik untuk menangkap pola periodik. Tahapan pengembangan model menggunakan pendekatan ensemble yang terbukti handal untuk data heterogen dengan ukuran terbatas. **XGBoost** dipilih karena kemampuannya menangani **interaksi fitur** yang kompleks dengan efisien dan performa superior pada dataset kecil (4), sementara **Random Forest** memberikan stabilitas prediksi dan kemampuan **generalisasi** yang baik (5). Kombinasi keduanya melalui pembobotan voting dengan bobot optimal yang ditentukan melalui set validasi diharapkan menghasilkan prediksi yang akurat dan tahan terhadap outlier.

**State of the Art dan Kebaruan**

Tabel 1 Analisis Komparatif Penelitian Prediksi Engagement Media Sosial

No	Penulis & Tahun	Metode	Dataset	Hasil	Kontribusi	Keterbatasan
1	Gorrepati (2024) (6)	XGBoost, Random Forest, BERT	Instagram; fokus klasifikasi <i>engagement</i>	Akurasi terbaik single fold dan stacked 89%	Bandingkan banyak algoritma + uji <i>word embeddings</i> & SMOTE untuk prediksi engagement	Bahasa Inggris; bukan jurnal
2	Podda & Recupero (2020) (7)	XGBoost dengan Apache Spark	Instagram, <b>106.404</b> baris/2.545 akun; <b>tanpa</b> analisis visual	<b>Balanced accuracy</b> terbaik $\approx$ <b>64.72%</b>	Formulasi <b>binary classification</b> (popular/unpopular) & <i>feature engineering</i> caption/emoji/hashtag	Tidak memanfaatkan fitur visual; metrik dominan balanced



Peta Jalan Penelitian 5 Tahun digambarkan pada Gambar 1. **Tahun 2025** fokus pada **pengembangan model dasar** dengan dataset 227 postingan FST UNJA menggunakan ensemble XGBoost-Random Forest, validasi komprehensif dengan validasi silang, dan publikasi hasil awal di konferensi nasional serta pengembangan aplikasi prototipe untuk penggunaan internal fakultas. **Tahun 2026** akan mengeksplorasi **deep learning** dengan transfer learning dari pre-trained models seperti ResNet untuk fitur visual dan IndoBERT untuk **fitur tekstual** guna meningkatkan performa dengan data terbatas, serta integrasi analisis Instagram Stories dan Reels yang memiliki karakteristik keterlibatan berbeda dengan ekspansi dataset melalui kolaborasi dengan fakultas lain di UNJA. **Tahun 2027** lanjut mengembangkan **mekanisme atensi** untuk memahami **interaksi antar fitur** secara lebih mendalam, **implementasi** sistem prediksi secara **real-time** dengan pembelajaran berkelanjutan dari umpan balik keterlibatan aktual, dan pengembangan API untuk integrasi dengan sistem manajemen konten fakultas.

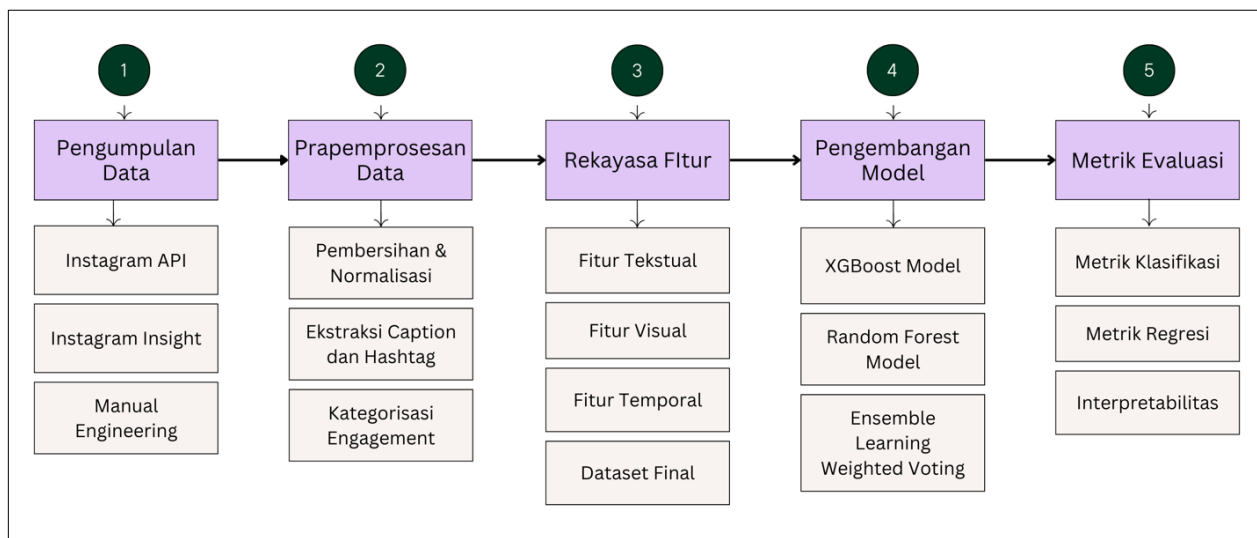
**Tahun 2028** mulai meneliti **kausalitas** menggunakan metode **inferensi kausal** untuk memahami **faktor penyebab keterlibatan** bukan hanya korelasi, pengembangan asisten **pembuatan konten berbasis AI** yang dapat menghasilkan caption optimal dan rekomendasi visual berdasarkan pembelajaran dari postingan sukses, serta kolaborasi dengan universitas lain di Sumatera untuk validasi lintas institusi. **Tahun 2029** melakukan **integrasi multi-platform** mencakup Instagram, TikTok, dan Twitter dengan model keterlibatan terpadu, **pengembangan framework open-source** untuk institusi pendidikan Indonesia dilengkapi dokumentasi komprehensif dan dukungan komunitas untuk adopsi nasional, serta publikasi di jurnal internasional bereputasi untuk diseminasi global.

## METODA

Metode atau cara untuk mencapai tujuan yang telah ditetapkan ditulis tidak melebihi 1200 kata. Bagian ini dapat dilengkapi dengan diagram alir penelitian yang menggambarkan apa yang sudah dilaksanakan dan yang akan dikerjakan selama waktu yang diusulkan. Format diagram alir dapat berupa file JPG/PNG. Metode penelitian harus dibuat secara utuh dengan penahapan yang jelas, mulai dari awal bagaimana proses dan luarannya, dan indikator capaian yang ditargetkan yang tercermin dalam Rencana Anggaran Biaya (RAB).

## Metode Penelitian

Penelitian akan menggunakan pendekatan **kuantitatif eksperimental** dengan desain pembelajaran mesin untuk **mengembangkan model prediksi keterlibatan** postingan Instagram. Metodologi dirancang dalam **lima tahap** sistematis mulai dari pengumpulan data, prapemrosesan, rekayasa fitur, pengembangan model, hingga evaluasi akhir yang terintegrasi untuk memastikan validitas dan aplikabilitas hasil dengan dataset terbatas namun representatif. Tahapan umum dan langkah nya digambarkan pada Gambar 2.



Gambar 2 Flowchart Keseluruhan Metode Penelitian yang diusulkan

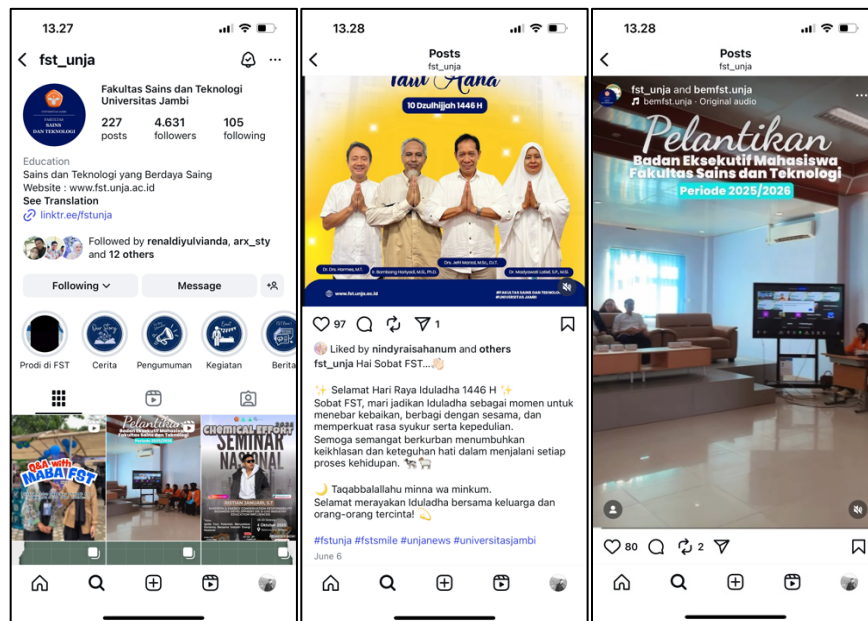
## Tahap 1: Pengumpulan Data

Pengumpulan data dilakukan dari akun Instagram @fst\_unja menggunakan **kombinasi Instagram Graph API dan ekstraksi manual** untuk memastikan kelengkapan dan akurasi data. Total 227 postingan yang telah dipublikasikan dikumpulkan secara menyeluruh mencakup **periode tiga tahun terakhir** dari Januari 2022 hingga Desember 2024. Periode ini dipilih **untuk menangkap variasi temporal** yang komprehensif termasuk fluktuasi musiman akibat kalender akademik. Instagram Graph API digunakan untuk mengakses metadata postingan secara sistematis dengan autentikasi OAuth 2.0, memungkinkan ekstraksi data dalam format JSON yang terstruktur (12), secara detail diperlihatkan Tabel 2.

Tabel 2 Jenis Data yang Dikumpulkan

Sumber	Kategori Data	Elemen Data	Format	Periode Pengambilan
Instagram API	Metadata Post	Caption, hashtag, timestamp, tipe media	JSON	Real-time via API
Instagram Insights	Metrik Keterlibatan	Likes, comments, shares, saves	CSV	7 hari pasca-publikasi
Instagram	Konten Visual	Gambar, video thumbnail	JPEG/MP4	Saat publikasi
Manual	Kalender Akademik	Periode ujian, wisuda, pendaftaran	Excel	Per semester
Manual	Konteks Acara	Kegiatan fakultas, hari libur nasional	CSV	Bulanan

**Data kontekstual** berupa **kalender akademik UNJA** dikumpulkan untuk memahami pengaruh periode akademik terhadap pola keterlibatan. Informasi tentang periode penerimaan mahasiswa baru, ujian tengah semester, ujian akhir semester, wisuda, dan libur akademik dipetakan ke setiap postingan menggunakan timestamp untuk identifikasi periode yang relevan. Pengumpulan data juga **mempertimbangkan event khusus** fakultas seperti seminar, workshop, kompetisi mahasiswa, dan kegiatan kemahasiswaan lainnya yang dapat mempengaruhi pola keterlibatan seperti pada Gambar 3.



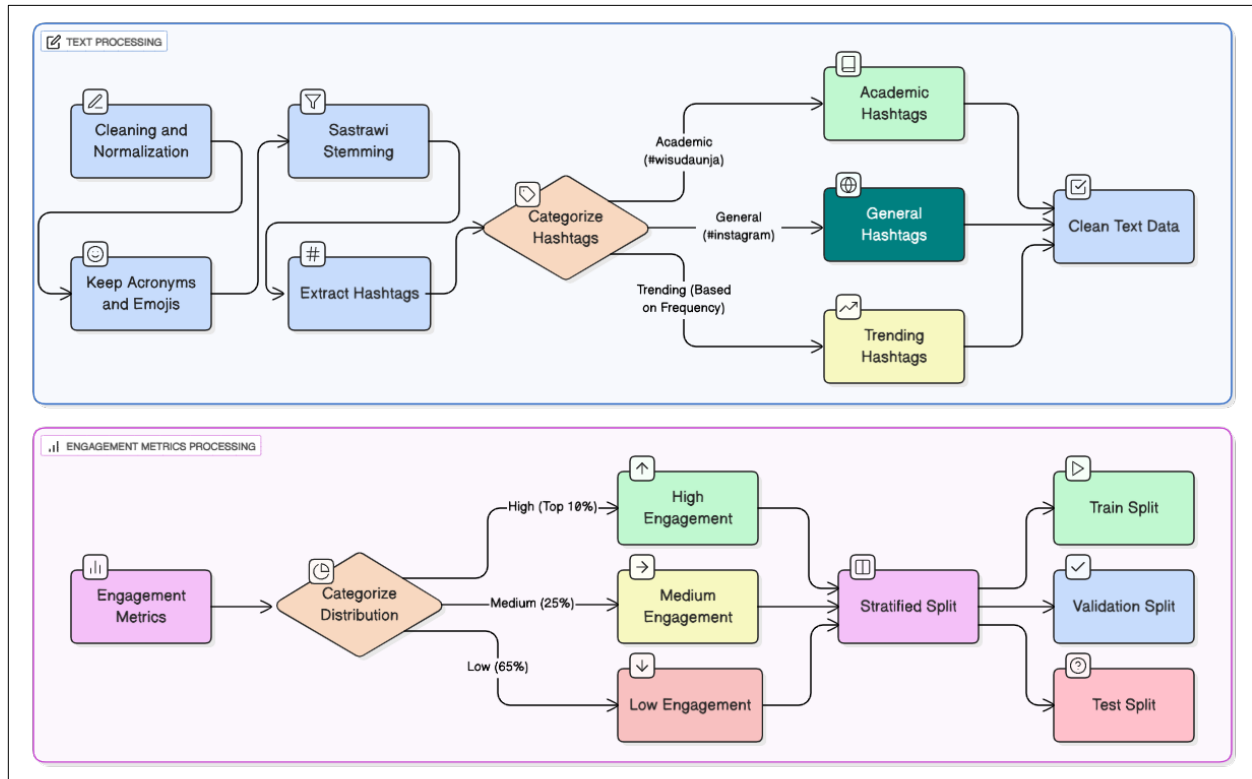
Gambar 3 Contoh Post dari akun instagram Fakultas Sains dan Teknologi @fst\_unja

## Tahap 2: Prapemrosesan Data

Prapemrosesan dimulai dengan **pembersihan dan normalisasi** data untuk **memastikan konsistensi dan kualitas** input model, secara lengkap dapat dilihat pada Gambar 4. Data tekstual dari caption dinormalisasi dengan konversi ke huruf kecil untuk konsistensi, namun mempertahankan kapitalisasi untuk akronim

institusional seperti FST, UNJA, dan nama program studi. Karakter khusus dihilangkan kecuali emoji yang informatif karena penelitian menunjukkan emoji memiliki nilai prediktif untuk keterlibatan. Koreksi typo dilakukan menggunakan kamus khusus yang dikembangkan untuk konteks akademik Indonesia.

Caption kemudian diproses menggunakan library Sastrawi untuk stemming dengan kamus tambahan berisi istilah akademik lokal yang tidak ada dalam kamus standar bahasa Indonesia (13). Proses ini penting untuk mengurangi variasi morfologis kata dalam bahasa Indonesia yang kaya akan imbuhan. Hashtag diekstrak dari caption dan **dikategorikan berdasarkan relevansi** dan popularitas menjadi tiga kelompok yaitu akademik spesifik seperti #wisudaunja, umum seperti #instagram, dan trending yang sedang berlangsung berdasarkan frekuensi historis.



Gambar 4 Tahapan Lengkap Prapemrosesan Data dari Raw Data ke Siap Diolah

**Kategorisasi keterlibatan** dilakukan berdasarkan distribusi data aktual untuk dapat melakukan proses evaluasi model dari perspektif **klasifikasi** dan **regresi**. Dataset dibagi **secara acak** dengan pengambilan **sampel berstrata** untuk mempertahankan **proporsi kelas** keterlibatan yaitu set pelatihan 70 persen atau 159 postingan, set validasi 15 persen atau 34 postingan, dan set pengujian 15 persen atau 34 postingan. Seed acak ditetapkan pada nilai 42 untuk memastikan penelitian dapat direproduksi.

### Tahap 3: Rekayasa Fitur

Rekayasa fitur dilakukan secara komprehensif pada tiga dimensi utama dengan total 55 fitur yang dirancang untuk menangkap aspek komprehensif dari setiap postingan Instagram seperti Tabel 3.

Tabel 3 Deskripsi Fitur yang Diekstrak

Dimensi	Kategori	Fitur	Jumlah	Deskripsi
Tekstual	Dasar	Panjang caption, jumlah kalimat, rata-rata kata	5	Statistik dasar teks
	Linguistik	Penggunaan emoji, mention, URL, pertanyaan	6	Elemen interaktif
	Sentimen	Polaritas, subjektivitas	2	Analisis sentimen Indonesia



	Hashtag	Jumlah, popularitas, keberagaman	5	Analisis hashtag historis
	TF-IDF	Vektor topik utama	7	Topik dominan
<b>Visual</b>	Warna	Mean & standar deviasi histogram RGB/HSV	12	Distribusi warna
	Deteksi	Jumlah wajah, flag keberadaan wajah	2	Haar Cascade
	Komposisi	Kecerahan, kontras, saturasi, tepi	4	Karakteristik visual
	Format	Tipe media, rasio aspek	2	Metadata gambar
<b>Temporal</b>	Waktu	Pengkodean siklik jam	2	Pengkodean siklik
	Hari	Pengkodean one-hot hari	7	Pola mingguan
	Konteks	Flag periode akademik	1	Kalender akademik

**Fitur tekstual** menggunakan pemrosesan bahasa alami untuk bahasa Indonesia dengan library Sastrawi dan analisis sentimen yang diadaptasi untuk konteks lokal. **Fitur visual** diproses menggunakan OpenCV untuk ekstraksi karakteristik gambar dengan algoritma yang telah dioptimasi untuk efisiensi komputasi (14). Deteksi wajah dilakukan menggunakan Haar Cascade classifier yang merupakan metode efisien untuk deteksi objek real-time (15). **Fitur temporal** menggunakan pengkodean siklik untuk mempertahankan kontinuitas waktu dan menghindari diskontinuitas pada batas periode.

#### Tahap 4: Pengembangan Model

Model dikembangkan menggunakan dua algoritma pembelajaran mesin yang telah terbukti efektif untuk prediksi dengan data tabular dengan konfigurasi seperti Tabel 4. Konfigurasi optimal ditentukan melalui pencarian grid dengan validasi silang 5 lipatan untuk menghindari overfitting pada dataset kecil (17).

Tabel 4 Konfigurasi Model dan Parameter

Model	Parameter	Nilai	Justifikasi
<b>XGBoost</b>	n_estimators	300	Optimal untuk dataset kecil
	max_depth	5	Mencegah overfitting
	learning_rate	0.1	Konvergensi stabil
	subsample	0.8	Menambah keacakan
	colsample_bytree	0.8	Sampling fitur
	scale_pos_weight	Auto	Mengatasi ketidakseimbangan kelas
<b>Random Forest</b>	n_estimators	200	Stabilitas prediksi
	max_depth	8	Kompleksitas moderat
	min_samples_split	5	Kontrol pemisahan
	min_samples_leaf	3	Ukuran daun minimum
	class_weight	balanced	Penyesuaian otomatis
<b>Ensemble</b>	Bobot XGBoost	0.6	Berdasarkan validasi
	Bobot RF	0.4	Komplemen stabilitas

**XGBoost** dipilih karena kemampuannya menangani interaksi fitur yang kompleks dengan efisien dan telah terbukti efektif pada berbagai task prediksi media sosial (4). Algoritma ini menggunakan gradient boosting

framework yang membangun pohon keputusan secara sekuensial dimana setiap pohon baru memperbaiki kesalahan pohon sebelumnya.

**Random Forest** memberikan stabilitas prediksi dan interpretabilitas yang baik melalui feature importance (5). Algoritma ini membangun multiple decision trees secara paralel dengan random sampling dan feature selection, kemudian mengagregasi prediksi melalui voting. Kombinasi keduanya melalui pembobotan voting menghasilkan prediksi yang lebih handal dibanding model tunggal dengan **bobot optimal XGBoost 0.6 dan Random Forest 0.4** yang ditentukan melalui validation set. Implementasi model menggunakan library Scikit-learn yang menyediakan interface untuk berbagai algoritma pembelajaran mesin (16)

## Tahap 5: Metrik Evaluasi

Evaluasi model dilakukan secara komprehensif menggunakan berbagai metrik untuk memahami performa dari berbagai perspektif yang daftarnya seperti pada Tabel 5. Evaluasi dilakukan pada set pengujian yang terpisah untuk memastikan generalisasi model.

Tabel 5 Metrik Evaluasi Model

Kategori	Metrik	Rumus	Deskripsi
Klasifikasi	Akurasi	$(TP+TN)/(TP+TN+FP+FN)$	Ketepatan keseluruhan
	Presisi	$TP/(TP+FP)$	Ketepatan prediksi positif
	Recall	$TP/(TP+FN)$	Sensitivitas deteksi
	F1-Score	$2 \times (\text{Presisi} \times \text{Recall}) / (\text{Presisi} + \text{Recall})$	Rata-rata harmonik
Regresi	MAE	$\sum  y_{\text{aktual}} - y_{\text{prediksi}}  / n$	Kesalahan absolut rata-rata
	RMSE	$\sqrt{(\sum (y_{\text{aktual}} - y_{\text{prediksi}})^2) / n}$	Kesalahan kuadrat rata-rata
	MAPE	$(100/n) \times \sum  (y_{\text{aktual}} - y_{\text{prediksi}}) / y_{\text{aktual}} $	Kesalahan persentase
Interpretabilitas	SHAP	Nilai Shapley	Kontribusi fitur
	Feature Importance	Gini/Gain	Signifikansi variabel

**Validasi silang dengan 5 lipatan** berstrata dilakukan untuk mengestimasi performa generalisasi dan mendeteksi overfitting (17). Untuk mengatasi ketidakseimbangan kelas, teknik SMOTE dapat diterapkan jika diperlukan (18). Learning curves dianalisis untuk memahami perilaku model terhadap ukuran dataset dan mengidentifikasi apakah model mengalami high bias atau high variance. Calibration plots digunakan untuk menilai keandalan probabilitas prediksi model.

**Nilai SHAP** digunakan untuk interpretabilitas model, memberikan wawasan tentang kontribusi setiap fitur terhadap prediksi baik secara global maupun lokal untuk setiap prediksi individual (11). **Analisis kesalahan** dilakukan untuk memahami pola kegagalan model. **Confusion matrix** digunakan untuk visualisasi kesalahan klasifikasi antar kelas. Analisis residual untuk aspek regresi membantu mengidentifikasi bias sistematis dalam prediksi. Nilai Feature importance yang dihasilkan kedua model dibandingkan untuk memvalidasi konsistensi faktor-faktor yang mempengaruhi keterlibatan. Evaluasi metrik menggunakan pendekatan komprehensif yang mencakup presisi, recall, dan F1 score macro dan micro untuk memberikan gambaran lengkap performa model (19).

Evaluasi juga mencakup analisis temporal untuk memahami performa model pada periode akademik yang berbeda. Hal ini penting mengingat pola keterlibatan yang berbeda antara periode aktif perkuliahan, ujian, dan libur akademik. Robustness testing juga dilakukan dengan simulasi berbagai skenario input untuk memastikan stabilitas prediksi model dalam kondisi operasional nyata. Adaptasi terhadap perubahan pola keterlibatan dari waktu ke waktu juga perlu dipertimbangkan untuk menjaga relevansi model (20).

Jadwal penelitian disusun berdasarkan pelaksanaan penelitian, harap disesuaikan berdasarkan lama tahun pelaksanaan penelitian

Penelitian dirancang untuk diselesaikan dalam 12 bulan dengan pendekatan berbasis pencapaian yang memungkinkan pemantauan kemajuan dan penyesuaian jika diperlukan.

No	Kegiatan	Bulan											
		1	2	3	4	5	6	7	8	9	10	11	12
	<b>A. Tahap Persiapan</b>												
1	Studi literatur dan tinjauan pustaka												
2	Persiapan infrastruktur teknis dan konfigurasi API												
3	Koordinasi dengan tim Humas FST												
	<b>B. Tahap Pengumpulan Data</b>												
4	Ekstraksi data Instagram via API												
5	Pengumpulan data kalender akademik												
6	Validasi dan pemeriksaan kualitas dataset												
	<b>C. Tahap Rekayasa Fitur</b>												
7	Ekstraksi fitur tekstual dan pemrosesan bahasa												
8	Ekstraksi fitur visual dengan computer vision												
9	Pengkodean fitur temporal dan integrasi												
	<b>D. Tahap Pengembangan Model</b>												
10	Pengembangan model dasar												
11	Pelatihan XGBoost dan Random Forest												
12	Optimasi ensemble dan penyetelan parameter												
	<b>E. Tahap Evaluasi dan Validasi</b>												
13	Validasi silang dan pengujian performa												
14	Analisis SHAP untuk interpretabilitas												
15	Pengujian penerimaan pengguna												
	<b>F. Tahap Implementasi dan Dokumentasi</b>												
16	Pengembangan aplikasi web Streamlit												
17	Integrasi sistem dan pengaturan API												
18	Dokumentasi teknis												
19	Penulisan artikel ilmiah												
20	Laporan akhir dan diseminasi												

DAFTAR PUSTAKA



Sitasi disusun dan ditulis berdasarkan sistem nomor sesuai dengan urutan pengutipan, mengikuti format Vancouver. Hanya pustaka yang disitasi pada usulan penelitian yang dicantumkan dalam Daftar Pustaka.

## Daftar Pustaka

1. Kemp S. Digital 2024: Global overview report. We Are Social and Kepios. 2024 Jan;15(1):1-304.
2. Asosiasi Penyelenggara Jasa Internet Indonesia. Laporan survei internet APJII 2023: Penetrasi dan perilaku pengguna internet Indonesia. APJII. 2023;8(1):1-146.
3. Li Y, Xie Y. Is a picture worth a thousand words? An empirical study of image content and social media engagement. *Journal of Marketing Research*. 2020;57(1):1-19.
4. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016:785-794.
5. Breiman L. Random forests. *Machine Learning*. 2001;45(1):5-32.
6. Gorrepati LP. Instagram data analysis using machine learning. Master's Projects. San Jose State University. 2024;1365. DOI: 10.31979/etd.jk9e-a23n.
7. Podda AS, Recupero DR. Popularity prediction of Instagram posts. *Information*. 2020;11(9):453.
8. De Kok S, Punt L, van den Puttelaar R, Ranta K, Schouten K, Frasincar F. Predicting the popularity of Instagram posts for a lifestyle magazine using deep learning. *2nd International Conference on Communication Systems, Computing and IT Applications*. 2017:174-177.
9. Arora A, Bansal S, Kandpal C, Aswani R, Dwivedi Y. Measuring social media influencer index-insights from Facebook, Twitter and Instagram. *Journal of Retailing and Consumer Services*. 2019;49:86-101.
10. Mazloom M, Rietveld R, Rudinac S, Worring M, Van Dolen W. Multimodal popularity prediction of brand-related social media posts. *Proceedings of the 24th ACM International Conference on Multimedia*. 2016:197-201.
11. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*. 2017;30:4765-4774.
12. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*. 2011;12:2825-2830.
13. Tala FZ. A study of stemming effects on information retrieval in Bahasa Indonesia. Institute for Logic, Language and Computation, Universiteit van Amsterdam. 2003.
14. Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2001;1:511-518.
15. Bradski G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*. 2000;25(11):120-125.
16. Instagram. Instagram Basic Display API Documentation. Meta for Developers. 2024. Available from: <https://developers.facebook.com/docs/instagram-basic-display-api>
17. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*. 1995;2:1137-1143.

18. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*. 2002;16:321-357.
19. Powers DM. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*. 2011;2(1):37-63.
20. Gama J, Žliobaitė I, Bifet A, Pechenizkiy M, Bouchachia A. A survey on concept drift adaptation. *ACM Computing Surveys*. 2014;46(4):1-37.