

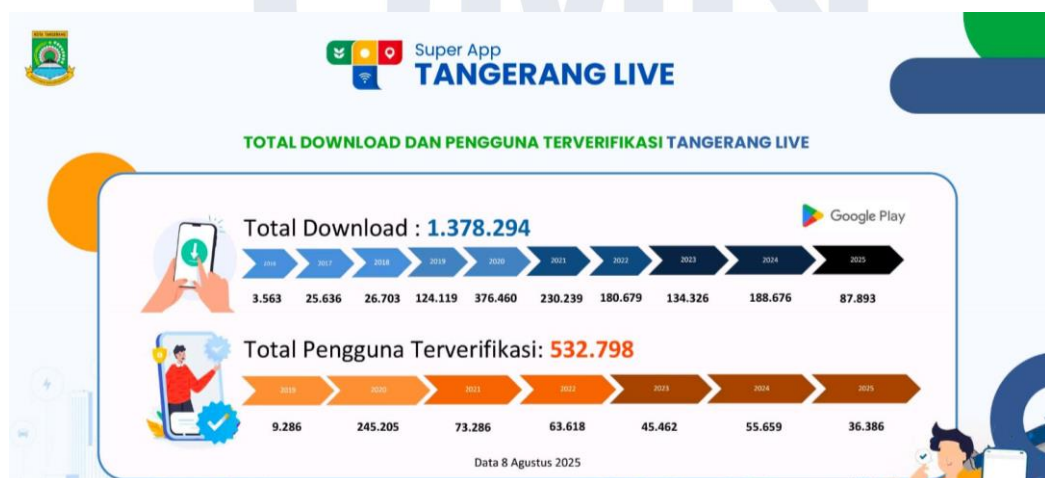
BAB I

PENDAHULUAN

1.1 Latar Belakang

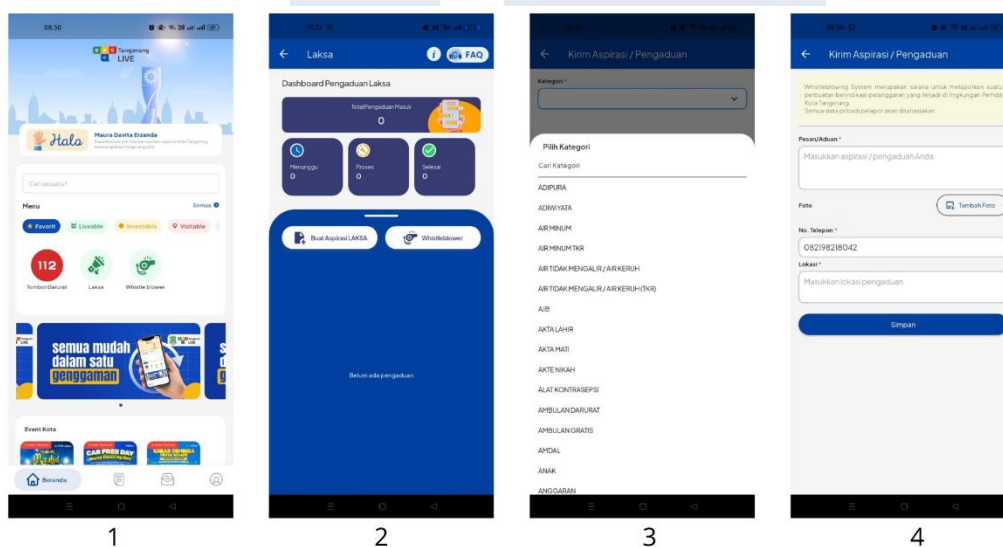
Transformasi digital menghasilkan Teknologi Informasi dan Komunikasi (TIK) yang memberikan pengaruh dalam layanan publik dengan berperan penting dalam meningkatkan kualitas pelayanan masyarakat [1]. Teknologi Informasi dan Komunikasi khususnya teknologi aplikasi sangat membantu dalam mengumpulkan data secara *real-time* dan dapat menganalisis data dalam waktu yang lebih singkat [2]. Proses analisis data yang lebih cepat akan membantu mempercepat pengambilan keputusan [3]. Proses pengambilan keputusan yang lebih cepat akan mempercepat penyelesaian masalah dan menciptakan hubungan baik antara pemerintah dan masyarakat [3].

Bentuk dukungan Pemerintah Kota Tangerang dalam transformasi digital diwujudkan dengan pembuatan aplikasi Tangerang LIVE. Aplikasi Tangerang LIVE merupakan hasil upaya pemerintah Kota Tangerang dalam mewujudkan pelayanan publik yang mudah, cepat, dan transparan. Tangerang LIVE disebut *super app* karena memiliki 12 layanan dan 34 menu untuk mendapatkan layanan administrasi kependudukan, laporan pengaduan masyarakat, dan pembayaran pajak daerah hingga zakat.



Gambar 1.1 Angka Pertumbuhan Aplikasi Tangerang LIVE

Gambar 1.1 yang bersumber dari *Google Play Store* yang diakses pada 9 Agustus 2025 menunjukkan aplikasi Tangerang Live mengalami pertumbuhan yang signifikan sejak peluncurannya. Diketahui bahwa *super app* Kota Tangerang ini pernah meraih posisi terbaik kedua dalam Indonesia's SDG's Action Awards 2024 atas pengembangan inovasi yang dilakukan [4]. Hingga bulan Agustus 2025, total pengunduhan aplikasi telah mencapai 1.378.294 dengan 532.798 pengguna terverifikasi. Salah satu pilar utama dari interaksi ini adalah fitur LAKSA (Layanan Aspirasi Kotak Saran Anda), yang dirancang khusus sebagai kanal utama bagi ratusan ribu pengguna tersebut untuk menyampaikan aspirasi, keluhan, dan saran secara langsung kepada pemerintah kota.



Gambar 1.2 langkah-langkah mengajukan pengaduan pada fitur LAKSA

Gambar 1.2 menunjukkan langkah-langkah yang harus dilakukan jika ingin mengajukan pengaduan pada fitur LAKSA aplikasi Tangerang LIVE. Langkah pertama dimulai dari gambar pertama yang menunjukkan halaman utama aplikasi Tangerang LIVE yang memiliki beberapa menu utama salah satunya LAKSA. Ketika menu LAKSA di pilih, pengguna akan memasuki halaman pada gambar kedua yang menunjukkan dashboard dengan *highlight box* yang menunjukkan total pengaduan yang masuk, laporan dengan status menunggu, di proses, dan selesai di proses. Jika *user* ingin membuat laporan, *user* bisa memilih opsi “Buat aspirasi LAKSA”. Gambar ketiga akan menunjukkan opsi kategori pengaduan yang ingin

diadukan dan kemudian memasuki halaman pada gambar keempat yang menunjukkan tampilan halaman yang memberikan kebebasan bagi pengguna untuk memasukkan laporan beserta gambarnya.

Fitur LAKSA yang memberikan kesempatan kepada masyarakat untuk mengajukan pengaduan menyebabkan besarnya volume pengaduan yang diasumsikan bahwa sebagian besar di antaranya termasuk dalam kategori negatif. Namun, tanpa adanya validasi atau data penunjang yang tepat, asumsi ini masih bersifat spekulasi. Oleh karena itu, penelitian ini dilakukan dengan tujuan memberikan bukti empiris terhadap data pengaduan yang terkumpul pada fitur LAKSA melalui analisis sentimen, yang diharapkan dapat menghasilkan informasi akurat mengenai proporsi sentimen positif dan negatif.

Hasil dari analisis sentimen yang dilakukan terhadap data dari fitur LAKSA akan diarahkan untuk mendukung proses pemantauan. Kemampuan memantau hasil sentimen akan membantu Diskominfo untuk mengetahui apakah pengaduan yang masuk termasuk dalam kategori negatif ataupun positif sebagai bahan evaluasi layanan publik. Manfaat dari hasil analisis ini juga memungkinkan Diskominfo untuk merespons isu-isu krusial yang muncul di masyarakat dengan lebih cepat dan proaktif.

Dalam upaya menjawab kebutuhan akan validasi data yang objektif dan monitoring pengaduan yang *real-time*, proses analisis sentimen akan dilakukan dengan pendekatan *Machine Learning* [5]. Analisis sentimen merupakan cabang dari *Natural Language Processing* (NLP) yang berfungsi mengidentifikasi dan mengklasifikasikan opini dalam teks [6]. Proses ini memungkinkan analisis puluhan ribu pengaduan secara konsisten dan terukur serta mengatasi analisis manual yang subjektif dan lambat [7]. Oleh karena itu, penting untuk melakukan sentimen analisis sebagai bukti empiris untuk memvalidasi kategori pengaduan yang masuk pada fitur LAKSA.

Dalam hal sentimen analisis, beberapa algoritma telah dikembangkan untuk membantu mengkategorikan sentiment [8]. Dua algoritma yang banyak digunakan dalam mengklasifikasikan sebuah sentimen diantaranya adalah *Support Vector*

Machine (SVM) dan *Random Forest* [9][10]. Faktor seperti informasi kontekstual, struktur kalimat, penggunaan bahasa formal atau informal, kata yang di singkat, dan konteks spesifik dari pengaduan layanan publik dapat memengaruhi hasil secara signifikan [11].

SVM bekerja dengan cara menemukan *hyperplane* optimal yang memisahkan titik data dari kelas yang berbeda dengan *margin* terbesar [11]. Keunggulan utama dari SVM terletak pada kemampuannya untuk bekerja secara efektif di ruang fitur berdimensi tinggi, serta kemampuannya dalam menangani data non-linear melalui teknik *kernel* [12]. Prinsip ini bergantung pada dua inti, yaitu maksimisasi *margin* dan *support vectors*, yang merupakan titik data yang paling dekat dengan batas keputusan tersebut [13]. Karena keandalannya dalam membangun model pemisah yang optimal, SVM telah terbukti menunjukkan akurasi tinggi dalam banyak penelitian maupun aplikasi, termasuk pengenalan wajah, diagnosis penyakit, pengenalan teks, dan keamanan jaringan [14].

Random Forest adalah metode pembelajaran ensemble yang kuat yang membangun beberapa pohon keputusan dan menggunakan suara mayoritas untuk prediksi akhir, secara efektif menangani dataset yang kompleks sambil mengurangi overfitting dibandingkan dengan pohon keputusan tunggal [12]. Namun, *Random Forest* tradisional menghadapi tantangan termasuk kurangnya keterjelasan dan masalah efisiensi komputasi seiring meningkatnya ukuran hutan [13]. Penelitian terbaru telah mengatasi keterbatasan ini melalui berbagai pendekatan seperti mengembangkan metode untuk menggabungkan *Random Forest* yang besar menjadi diagram keputusan tunggal, memberikan klasifikasi yang dapat dijelaskan dan mencapai percepatan beberapa urutan magnitudo dan memperkenalkan WildWood, yang meningkatkan prediksi dengan menggabungkan semua *subtree* yang mungkin menggunakan bobot eksponensial yang dihitung pada sampel *out-of-bag* [13].

Meskipun SVM dan *Random Forest* (RF) merupakan algoritma *machine learning* yang sangat kuat, keduanya memiliki pendekatan klasifikasi yang fundamental berbeda [14]. SVM bekerja dengan mencari *hyperplane* optimal di ruang berdimensi tinggi, sementara RF bekerja dengan membangun konsensus dari

sejumlah besar *decision tree* [14][15]. Perbandingan antara keduanya penting untuk menentukan model tunggal terbaik. Namun, penelitian ini berhipotesis bahwa performa yang lebih optimal dapat dicapai dengan mengembangkan model hibrida (*hybrid model*). Pendekatan hibrida ini bertujuan untuk menggabungkan kekuatan unik dari kedua algoritma tersebut, dengan harapan dapat menutupi kelemahan masing-masing model murni dan menghasilkan tingkat akurasi klasifikasi yang lebih tinggi untuk data pengaduan fitur LAKSA.

Hasil dari penelitian analisis sentimen pada fitur LAKSA aplikasi Tangerang LIVE diharapkan dapat memberikan kontribusi praktis bagi Dinas Komunikasi dan Informatika (Diskominfo) Kota Tangerang. Dengan penemuan model klasifikasi yang paling akurat, Diskominfo akan terbantu dalam memahami dan memvalidasi secara objektif opini masyarakat. Kontribusi utama dari penelitian ini akan diwujudkan dalam bentuk rekomendasi arsitektur model terbaik yakni model SVM, *Random Forest*, ataupun model hibrida SVM-RF yang terbukti paling unggul. Bagi masyarakat, penelitian ini secara tidak langsung akan meningkatkan kualitas respons pemerintah, sehingga keluhan dan masukan dapat dianalisis dan ditindaklanjuti dengan lebih efektif untuk mewujudkan visi *Smart City*.

1.2 Rumusan Masalah

Asumsi banyaknya pengaduan yang diterima berkonotasi negatif pada fitur LAKSA aplikasi Tangerang LIVE menjadi awal permasalahan yang dihadapi Dinas Komunikasi dan Informatika (Diskominfo) kota Tangerang, sehingga hal-hal yang menjadi rumusan masalah dalam penelitian ini diantaranya :

1. Bagaimana teknik *machine learning* dapat diterapkan untuk mengklasifikasikan sentimen pada data pengaduan dari fitur LAKSA di aplikasi Tangerang Live?
2. Bagaimana hasil perbandingan kinerja algoritma *support vector machine* (SVM) dan *Random Forest* dalam mengklasifikasikan sentimen pengaduan pada fitur LAKSA?

3. Bagaimana kinerja model hibrida (*hybrid*) yang menggabungkan *Support Vector Machine* (SVM) dan *Random Forest*, dan apakah kinerjanya mampu melampaui akurasi dari kedua model murni tersebut?

1.3 Batasan Masalah

Dalam upaya memastikan bahwa penelitian ini tetap fokus, mendalam, dan relevan dengan tujuan yang telah ditetapkan, maka ditentukan beberapa batasan masalah yang akan menjadi koridor pelaksanaan penelitian sebagai berikut :

1. **Objek penelitian dari fitur LAKSA Aplikasi Tangerang LIVE**

Penelitian ini hanya akan menggunakan data pengaduan yang bersumber secara eksklusif dari aplikasi resmi Pemerintah Kota Tangerang, yaitu fitur LAKSA pada aplikasi Tangerang LIVE. Keputusan ini diambil untuk menjamin keaslian dan konsistensi data, serta memastikan bahwa hasil analisis relevan secara langsung dengan kebutuhan Dinas Komunikasi dan Informatika (Diskominfo) selaku pengelola aplikasi. Penelitian tidak akan mencakup data dari *platform* lain seperti media sosial atau aplikasi tidak resmi.

2. **Batasan pada periode data yang digunakan**

Data yang akan digunakan dalam proses analisis terbatas pada periode waktu tertentu. Dalam analisis sentimen aplikasi Tangerang LIVE, data yang digunakan merupakan data pengaduan yang masuk selama periode tahun 1 Januari 2021 hingga 30 November 2024. Pembatasan waktu ini bertujuan untuk memastikan bahwa analisis yang dilakukan mencerminkan sentimen dan isu-isu terkini yang masih relevan dengan kondisi masyarakat saat ini. Mengingat aplikasi Tangerang LIVE yang telah dirilis sejak tahun 2016, pengambilan data empat tahun terakhir dilakukan untuk meminimalisir keterlibatan isu-isu lama yang tidak lagi relevan, sehingga rekomendasi yang diberikan jauh lebih akurat dan kontekstual terhadap permasalahan terkini.

3. **Batasan pada kelas klasifikasi sentimen**

Model klasifikasi yang akan dibangun hanya akan mengkategorikan sentimen ke dalam tiga kelas, yaitu positif, negatif, dan netral. Pembatasan

ini selaras dengan tujuan utama penelitian untuk memvalidasi asumsi dan menjawab pertanyaan utama Diskominfo mengenai polaritas dominan dari pengaduan yang masuk, serta menyederhanakan proses pemodelan agar lebih fokus dan efektif.

4. Batasan pada penggunaan algoritma dan evaluasi yang digunakan

Metode pemodelan dalam penelitian ini dibatasi pada implementasi dua algoritma pembelajaran mesin tunggal, yaitu *Support Vector Machine* (SVM) dan *Random Forest*, serta satu arsitektur model *hybrid* melalui teknik *Stacking*. Analisis yang dilakukan berfokus pada perbandingan performa antara model murni (*baseline*) dengan model *hybrid* guna mengetahui efektivitas penggabungan algoritma dalam meningkatkan akurasi klasifikasi. Penelitian ini tidak mencakup optimasi parameter secara luas pada algoritma di luar ketiga model tersebut atau penggunaan teknik *deep learning* lainnya. Sebagai instrumen evaluasi, penelitian ini menerapkan *confusion matrix* untuk mengukur kinerja model secara mendalam. Penggunaan *matrix* ini sangat krusial karena tidak hanya menyajikan angka akurasi global, tetapi juga mampu memberikan gambaran biner mengenai perbandingan antara kategorisasi aktual dan hasil prediksi algoritma. Melalui *matrix* ini, peneliti dapat mengidentifikasi serta menguantifikasi jumlah misklasifikasi pada setiap label sentimen, sehingga efektivitas model dalam menangani data pengaduan yang tidak seimbang dapat dianalisis secara objektif.

1.4 Tujuan dan Manfaat Penelitian

Analisis sentimen pada fitur LAKSA aplikasi Tangerang LIVE dilakukan untuk menerapkan teknik *machine learning* guna mengklasifikasikan sentimen pada pengaduan aplikasi Tangerang Live secara akurat. Hasil dari penelitian ini diharapkan dapat menjadi dasar bagi Diskominfo Kota Tangerang dalam merumuskan strategi peningkatan layanan publik yang lebih responsif dan berbasis data.

1.4.1 Tujuan Penelitian

Penelitian analisis sentimen pada fitur LAKSA aplikasi Tangerang LIVE memiliki tujuan yang diuraikan sebagai berikut :

1. **Menerapkan teknik *machine learning* untuk membangun model klasifikasi sentimen pada data pengaduan dari fitur LAKSA di aplikasi Tangerang Live**

Penelitian ini bertujuan untuk merancang sebuah sistem yang mampu mengolah data teks mentah dari pengaduan pengguna dan mengklasifikasikannya secara otomatis ke dalam kategori sentimen positif atau negatif. Hal ini akan menghasilkan sebuah metode analisis yang objektif dan terukur, menggantikan asumsi subjektif yang ada saat ini.

2. **Membandingkan efektivitas dan akurasi antara algoritma *Support Vector Machine* (SVM) dan *Random Forest* dalam mengklasifikasikan pengaduan**

Melalui analisis komparatif terhadap hasil klasifikasi kedua algoritma, penelitian ini dilakukan dengan tujuan menelaah detail performa masing-masing model. Telaah performa masing-masing model berguna dalam mengidentifikasi algoritma mana yang terbukti terbaik dan mampu menangani karakteristik data pengaduan pada fitur LAKSA secara optimal.

3. **Merancang dan mengevaluasi model hibrida (*hybrid*) SVM-*Random Forest* untuk optimasi akurasi klasifikasi**

Setelah mengevaluasi kinerja model SVM dan *Random Forest* sendiri, penelitian analisis sentimen yang dilakukan bertujuan untuk merancang dan membangun sebuah model hibrida yang mengkombinasikan kekuatan dari SVM dan *Random Forest*. Tujuan utama dari perancangan model hibrida adalah untuk menguji secara empiris apakah pendekatan hibrida ini mampu menghasilkan tingkat akurasi klasifikasi yang lebih superior dan optimal dibandingkan dengan hanya menggunakan SVM atau *Random Forest* secara terpisah.

1.4.2 Manfaat Penelitian

Penelitian yang menganalisis sentimen pada fitur LAKSA aplikasi Tangerang LIVE memiliki manfaat yang ditinjau dari dua aspek yang diuraikan sebagai berikut :

1. Manfaat Akademis

Kontribusi bagi pengembangan ilmu pengetahuan di bidang *Natural Language Processing* (NLP) dan *machine learning* akan diberikan melalui penelitian ini. Secara spesifik, literatur studi kasus akan diperkaya oleh adanya analisis komparatif antara model murni (*Support Vector Machine* dan *Random Forest*) dengan model hibrida (*hybrid model*). Sebuah dataset unik, yaitu teks pengaduan layanan publik berbahasa Indonesia dari super-app pemerintah, akan dimanfaatkan dalam studi ini. Hasil dari penelitian ini diharapkan dapat dijadikan referensi bagi peneliti selanjutnya yang akan melakukan analisis serupa, khususnya dalam domain *e-Government* atau optimalisasi model klasifikasi.

2. Manfaat Praktis

Secara praktis, manfaat langsung akan diberikan bagi Dinas Komunikasi dan Informatika (Diskominfo) Kota Tangerang. Melalui penelitian ini, sebuah rekomendasi arsitektur model (murni atau hibrida) dengan akurasi tertinggi akan dihasilkan sebagai dasar untuk validasi sentimen berbasis bukti. Dengan demikian, pemahaman objektif mengenai sentimen publik dapat diperoleh oleh Diskominfo, dan keputusan strategis dapat diambil secara lebih cepat. Bagi masyarakat Kota Tangerang, efektivitas penanganan keluhan diharapkan dapat ditingkatkan melalui temuan ini, sehingga kepuasan dan kepercayaan terhadap layanan pemerintah dapat turut ditingkatkan dalam mewujudkan visi Smart City.

1.5 Sistematika Penulisan

BAB 1 : PENDAHULUAN

Bab ini membahas latar belakang penelitian yang diawali dari tuntutan yang dihadapi pemerintah pada revolusi Industri 4.0 hingga pengenalan aplikasi Tangerang Live beserta fitur LAKSA sebagai objek utama. Dalam bab ini,

dirumuskan masalah utama yaitu kebutuhan untuk memvalidasi sentimen pengaduan secara objektif. Bab ini juga menjabarkan secara rinci rumusan masalah, tujuan penelitian yang ingin dicapai, serta manfaat akademis dan praktis yang diharapkan. Untuk menjaga fokus penelitian, ditetapkan pula batasan-batasan masalah yang jelas. Terakhir, sistematika penulisan dijelaskan untuk memberikan gambaran mengenai struktur dan alur keseluruhan laporan penelitian.

BAB 2 : LANDASAN TEORI

Bab ini menyajikan tinjauan pustaka dan kerangka teori yang menjadi fondasi penelitian. Diawali dengan pemaparan 10 penelitian terdahulu yang relevan untuk memposisikan penelitian ini di tengah riset yang sudah ada. Selanjutnya, dibahas teori-teori konseptual yang melingkupi topik skripsi, seperti Smart City, e-Government, Super App, Analisis Sentimen, *Natural Language Processing* (NLP), dan *Machine Learning*. Bab ini juga menjelaskan secara detail teori mengenai kerangka kerja CRISP-DM, algoritma *Random Forest*, *Support Vector Machine* (SVM), serta konsep inti dari arsitektur model hibrida (*hybrid model*) yang menjadi fokus optimasi. Sebagai penutup, dijelaskan pula fungsi dari perangkat lunak yang digunakan, seperti *Python*, *Visual Studio Code*, dan *Jupyter Notebook*.

BAB 3 : METODOLOGI

Bab ini menguraikan secara rinci langkah-langkah metodologis yang diterapkan dalam penelitian. Dimulai dengan gambaran umum objek penelitian yang berfokus pada analisis sentimen pengaduan pada fitur LAKSA di aplikasi Tangerang Live. Metode penelitian utama yang digunakan adalah kerangka kerja CRISP-DM, yang alurnya dijelaskan secara sistematis. Tahapan ini mencakup *Business Understanding*, *Data Understanding*, dan *Data Preparation* (termasuk pembersihan dan pra-pemrosesan teks). Tahap *Modeling* akan merinci pembangunan tiga arsitektur model: SVM murni (dengan optimasi *kernel*), *Random Forest* murni, dan model hibrida SVM-RF. Tahap *Evaluation* akan menjelaskan

perbandingan metrik performa dari ketiga model tersebut. Terakhir, fase *Deployment* diwujudkan dalam bentuk rekomendasi algoritma terbaik berdasarkan temuan evaluasi. Bab ini bertujuan untuk memberikan panduan yang transparan dan dapat direplikasi mengenai keseluruhan proses teknis yang dilakukan.

BAB 4 : ANALISIS DAN HASIL PENELITIAN

Bab ini merupakan inti dari penelitian yang menyajikan seluruh hasil implementasi teknis dari metodologi CRISP-DM dan pembahasannya secara mendalam. Bab ini akan merinci keseluruhan proses pada fase *Data Preparation*, mulai dari pembersihan data mentah, pra-pemrosesan teks (termasuk *stemming* Sastrawi), hingga proses *pseudo-labeling* yang menghasilkan 9.973 data bersih siap latih. Selanjutnya, bab ini akan memaparkan dan menganalisis secara rinci hasil dari fase *Modeling* dan *Evaluation*, yang mencakup perbandingan kinerja (termasuk *Classification Report* dan *Confusion Matrix*) dari model murni *Random Forest*, optimasi *kernel* pada model SVM, dan model hibrida SVM-RF. Pembahasan akan difokuskan pada analisis temuan kunci, pembuktian keunggulan model hibrida, dan diakhiri dengan fase *Deployment* yang merumuskan rekomendasi algoritma terbaik.

BAB 5 : SIMPULAN DAN SARAN

Bab ini merupakan bagian penutup dari keseluruhan laporan penelitian, yang terdiri dari Kesimpulan dan Saran. Bagian Kesimpulan akan berisi rangkuman jawaban atas tiga rumusan masalah, menegaskan hasil akhir perbandingan kinerja, dan menyatakan bahwa model hibrida (dengan akurasi 90.13%) terbukti sebagai arsitektur yang paling optimal. Bagian Saran akan memberikan rekomendasi praktis yang ditujukan kepada Diskominfo berdasarkan temuan tersebut, sekaligus memberikan saran akademis yang konstruktif untuk peneliti selanjutnya (seperti penggunaan *manual labeling* atau eksplorasi *deep learning*).