

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Era digitalisasi menuntut pemerintah daerah untuk memberikan layanan informasi publik yang cepat, akurat, dan mudah diakses oleh masyarakat luas. Kabupaten Rokan Hulu, yang dikenal dengan julukan "Negeri Seribu Suluk", memiliki potensi besar di berbagai sektor, mulai dari pariwisata, kebudayaan, tata kelola pemerintahan, hingga layanan publik administrasi. Penyebaran informasi yang transparan dan interaktif tidak hanya krusial untuk mendukung program *e-government*, tetapi juga untuk meningkatkan partisipasi masyarakat serta daya tarik kawasan bagi wisatawan luar daerah [1].

Saat ini, penyediaan informasi mengenai Kabupaten Rokan Hulu umumnya masih bertumpu pada portal situs web resmi yang bersifat statis atau layanan pelanggan secara manual. Pendekatan konvensional ini memiliki sejumlah keterbatasan operasional, seperti keharusan pengguna untuk mencari data secara manual di antara tumpukan artikel, waktu tunggu (*response time*) yang lambat pada layanan aduan, serta tidak adanya layanan di luar jam kerja staf operasional. Kendala-kendala ini sering kali menyebabkan masyarakat kesulitan mendapatkan jawaban spesifik secara instan atau *real-time* [2].

Untuk mengatasi keterbatasan aksesibilitas informasi tersebut, implementasi kecerdasan buatan (*Artificial Intelligence*) dalam bentuk *chatbot* interaktif menawarkan solusi inovatif yang sangat relevan. Pengembangan *Intelligent*

Regional Bot (IRB) memungkinkan terbentuknya asisten virtual cerdas yang beroperasi selama 24 jam penuh untuk melayani pertanyaan pengguna secara otomatis. Kehadiran sistem cerdas ini dapat mengubah paradigma pencarian informasi dari metode navigasi *website* yang membosankan menjadi interaksi percakapan dua arah yang jauh lebih praktis dan intuitif [3].

Meskipun demikian, sebagian besar *chatbot* konvensional yang digunakan pada layanan publik saat ini masih menggunakan metode *rule-based* atau pencocokan kata kunci (*keyword matching*). Sistem ini memiliki kelemahan yang sangat fatal karena bersifat kaku; bot hanya mampu merespons jika kalimat pengguna sama persis dengan aturan di dalam pangkalan data. Ketika masyarakat bertanya menggunakan struktur kalimat yang dinamis, tidak baku, atau mengandung kesalahan ketik (*typo*), *chatbot* konvensional umumnya gagal memahami maksud pengguna dan memberikan jawaban yang tidak relevan [4].

Menghadapi kompleksitas dan variasi bahasa sehari-hari yang digunakan masyarakat, pendekatan *Natural Language Processing (NLP)* berbasis *deep learning* menjadi kebutuhan mutlak. Salah satu arsitektur mutakhir yang terbukti memiliki performa superior dalam memproses bahasa Indonesia adalah model IndoBERT. Sebagai varian dari *Bidirectional Encoder Representations from Transformers (BERT)* yang telah dilatih secara khusus dengan jutaan korpus teks berbahasa Indonesia, *IndoBERT* memiliki kemampuan luar biasa dalam memahami makna dan konteks kalimat secara dua arah, menjadikannya jauh lebih cerdas daripada sekadar mengenali kata per kata [5].

Keunggulan model ini juga telah dibuktikan oleh sejumlah penelitian terbaru, salah satunya oleh Hidayat, dkk. (2024), yang menegaskan bahwa penerapan fine-tuning pada model *IndoBERT* untuk sistem tanya jawab (*Question Answering*) dan klasifikasi niat (*intent classification*) menghasilkan tingkat akurasi yang jauh lebih tinggi dan tahan terhadap variasi tata bahasa yang tidak terstruktur. Pemrosesan bahasa alami yang mendalam ini memungkinkan bot merespons pertanyaan kompleks seputar profil wilayah, prosedur administrasi, hingga rekomendasi wisata dengan tingkat pemahaman yang menyerupai manusia [6].

Merujuk pada urgensi ketersediaan layanan informasi publik yang responsif dan keunggulan teknologi pemrosesan bahasa alami tersebut, penelitian ini difokuskan pada perancangan sebuah *bot* cerdas untuk layanan informasi daerah. Sistem ini dibangun dengan memanfaatkan arsitektur *NLP* termutakhir untuk memahami dan memproses pertanyaan warga maupun wisatawan secara otomatis dengan tingkat akurasi tinggi [7]. Oleh karena itu, penelitian ini diangkat dengan judul **“Pengembangan IRB (*Intelligent Regional Bot*) Menggunakan model IndoBERT untuk informasi kabupaten rokan hulu”**.

1.2 Rumusan Masalah

Bagaimana merancang dan membangun aplikasi *Intelligent Regional Bot (IRB)* berbasis model *IndoBERT* yang mampu memahami dan menyajikan informasi seputar Kabupaten Rokan Hulu secara otomatis dan interaktif?

1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk merancang dan membangun aplikasi model pemrosesan bahasa alami *IndoBERT* guna mengembangkan aplikasi *Intelligent Regional Bot (IRB)* sebagai pusat layanan cerdas Kabupaten Rokan Hulu, sehingga dapat mempermudah masyarakat dan wisatawan dalam mengakses informasi daerah secara interaktif, cepat, dan otomatis.

1.4 Batasan Penelitian

Dalam penelitian ini, peneliti memberikan beberapa batasan penelitian yaitu:

1. *Intelligent Regional Bot (IRB)* hanya dirancang untuk memberikan informasi umum seputar Kabupaten Rokan Hulu, yang difokuskan pada bidang pariwisata, profil daerah, dan panduan layanan publik.
2. Sumber data (dataset) yang digunakan untuk melatih bot berasal dari situs web resmi pemerintah daerah, dokumen publik, serta artikel atau literatur terkait yang valid mengenai Kabupaten Rokan Hulu.
3. Sistem tanya jawab cerdas ini hanya memproses masukan berupa teks (berbahasa Indonesia) dan memberikan keluaran berupa teks (*text-to-text*), serta belum mendukung masukan berupa suara (*voice*) maupun gambar (*image*).
4. *Bot* ini bersifat sebagai asisten virtual pemberi informasi searah dan tidak terintegrasi langsung untuk memproses transaksi administrasi kependudukan yang bersifat rahasia secara otomatis.

5. Pembuatan model *Natural Language Processing (NLP)* menggunakan model arsitektur *IndoBERT*, dan bahasa pemrograman yang dipakai dalam pengembangan sistem ini adalah *Python*.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah:

1. Memudahkan masyarakat dan wisatawan dalam mengakses informasi terkait Kabupaten Rokan Hulu (seperti pariwisata, layanan publik, dan profil daerah) secara cepat, efisien, dan interaktif tanpa harus mencari secara manual atau datang langsung ke instansi pemerintah daerah.
2. Menjadi model penerapan *Intelligent Regional Bot (IRB)* berbasis AI dan pemrosesan bahasa alami (*IndoBERT*) dalam bidang pelayanan informasi publik, yang dapat dikembangkan lebih lanjut untuk mendukung sistem *e-government* di daerah lain.
3. Meningkatkan eksposur, daya tarik wisata, dan pemahaman masyarakat luas terhadap potensi Kabupaten Rokan Hulu melalui pendekatan teknologi komunikasi yang modern dan mudah diakses.

1.5 Sistematika Penulisan

BAB 1 PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, tujuan penelitian, batasan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB 2 LANDASAN TEORI

Bab ini berisi teori – teori yang digunakan pada penelitian ini. Teori – teori yang berhubungan dengan Perancangan Aplikasi Pengembangan *IRB (Intelligent Regional Bot)* Menggunakan Model *IndoBERT* Untuk Informasi Kabupaten Rokan Hulu.

BAB 3 METODOLOGI PENELITIAN

Bab ini berisi tahapan-tahapan penelitian, mulai dari identifikasi masalah, pengumpulan data, analisa model, perancangan sistem, implementasi sistem, dan pengujian sistem.

BAB 4 ANALISIS DAN PERANCANGAN SISTEM

Bab ini berisi Analisa dan Perancangan Aplikasi Pengembangan *IRB (Intelligent Regional Bot)* Menggunakan Model *IndoBERT* Untuk Informasi Kabupaten Rokan Hulu.

BAB 5 IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi implementasi dan pengujian pada aplikasi yang akan dibangun.

BAB 6 PENUTUP

Bab ini berisi kesimpulan dari penelitian yang telah dilakukan serta saran untuk penelitian atau pengembangan sistem di masa mendatang.

BAB 2

LANDASAN TEORI

2.1 Informasi Publik dan E-Government

Perkembangan teknologi informasi dan komunikasi telah mendorong transformasi besar dalam tata kelola pemerintahan di berbagai belahan dunia, termasuk Indonesia. Fenomena ini melahirkan konsep *e-government*, yaitu pemanfaatan teknologi digital oleh lembaga pemerintah untuk meningkatkan efisiensi, transparansi, akuntabilitas, serta efektivitas pemberian layanan kepada masyarakat. Melalui penerapan *e-government*, paradigma pelayanan publik bergeser dari yang semula bersifat birokratis dan kaku menjadi lebih dinamis serta berorientasi pada pemenuhan kebutuhan warga secara cepat dan tepat [8].

Salah satu pilar utama dalam keberhasilan *e-government* adalah penyediaan layanan informasi publik yang inklusif dan mudah diakses. Pemerintah daerah berkewajiban untuk menyebarluaskan data mengenai potensi wilayah, regulasi, prosedur administratif, hingga program kerja kepada masyarakat secara transparan. Keterbukaan informasi ini bukan hanya sekadar pemenuhan kewajiban hukum, melainkan instrumen krusial untuk membangun kepercayaan publik (*public trust*) serta mendorong partisipasi aktif masyarakat dalam pembangunan daerah [9].

Namun, implementasi layanan informasi digital di tingkat regional sering kali menghadapi kendala keterbatasan interaksi. Banyak portal informasi pemerintah daerah yang masih mengandalkan platform *web* statis, di mana pengguna harus

menyisir tumpukan dokumen digital secara manual untuk menemukan informasi spesifik. Oleh karena itu, diperlukan inovasi teknologi yang mampu mentransformasikan platform statis tersebut menjadi sistem layanan informasi interaktif yang dapat merespons kebutuhan masyarakat secara real-time tanpa batasan waktu dan ruang [10].

2.2 Artificial Intellignce dan Chatbot

Artificial Intelligence (AI) atau kecerdasan buatan merupakan salah satu cabang ilmu komputer yang berfokus pada pengembangan sistem komputer agar mampu melakukan tugas-tugas yang biasanya memerlukan kecerdasan manusia. Tugas-tugas tersebut meliputi kemampuan untuk belajar dari pengalaman, bernalar untuk mengambil keputusan, memecahkan masalah kompleks, serta memahami bahasa manusia. Di sektor pelayanan publik, implementasi *AI* kini diarahkan untuk mengotomatisasi interaksi antara instansi penyedia layanan dengan masyarakat secara massal namun tetap personal [11].

Salah satu manifestasi praktis dari teknologi *AI* yang paling populer dalam layanan pelanggan adalah chatbot atau agen virtual. Chatbot merupakan sebuah program komputer yang dirancang untuk mensimulasikan percakapan intelektual dengan satu atau lebih pengguna manusia, baik melalui media teks maupun suara. Dalam ekosistem digital modern, *chatbot* berperan penting sebagai lini terdepan pelayanan yang mampu menangani ribuan pertanyaan secara serentak, mengurangi beban kerja staf operasional manusia, serta menekan biaya operasional instansi secara signifikan [12].

Seiring berjalannya waktu, teknologi *chatbot* mengalami evolusi yang sangat pesat. Sistem *chatbot* generasi awal umumnya dibangun berbasis aturan kaku (*rule-based*), yang hanya mengandalkan pencocokan kata kunci (*keyword matching*) atau pohon keputusan sederhana untuk merespons pengguna. Generasi *chatbot* masa kini telah beralih ke pendekatan berbasis kecerdasan buatan komputasional yang adaptif, di mana sistem tidak lagi sekadar mencari kesamaan kata, melainkan mampu menganalisis pola kalimat, memahami intensi tersirat, dan memberikan jawaban yang lebih kontekstual serta natural [13].

2.3 Natural Language Processing (NLP)

Natural Language Processing (NLP) atau Pemrosesan Bahasa Alami adalah sub-bidang dari kecerdasan buatan dan linguistik komputasional yang menjembatani interaksi antara komputer dan bahasa manusia. Fokus utama dari *NLP* adalah memberikan kemampuan kepada komputer untuk membaca, menguraikan, memahami, dan mengekstrak makna dari bahasa manusia dengan cara yang bernilai. Tantangan terbesar dalam *NLP* terletak pada sifat bahasa manusia yang sangat kompleks, penuh dengan ambiguitas, metafora, dialek daerah, serta struktur kalimat yang tidak baku dalam komunikasi sehari-hari [14].

Dalam arsitektur sistem informasi cerdas, *NLP* bertindak sebagai mesin penerjemah yang mengubah teks bebas (*unstructured text*) dari pengguna menjadi representasi data terstruktur yang dapat dipahami oleh algoritma komputer. Proses ini melibatkan berbagai tahapan prapengolahan data teks (*text preprocessing*), seperti pembersihan teks (*cleansing*), pemisahan kata (*tokenization*), penghapusan kata sambung (*stopword removal*), hingga pengubahan kata berimbuhan menjadi

kata dasar (*stemming* atau *lemmatization*). Keberhasilan tahapan awal ini sangat menentukan kualitas fitur teks yang akan dianalisis pada tahap pemodelan selanjutnya [15].

Kini, pemanfaatan *NLP* telah meluas ke berbagai aplikasi praktis, mulai dari analisis sentimen, penerjemahan bahasa otomatis, ekstraksi informasi, hingga pembangunan sistem tanya jawab interaktif. Dalam konteks pengembangan asisten virtual daerah, teknologi *NLP* memegang peran sentral untuk membedah setiap pertanyaan yang diajukan oleh masyarakat. Tanpa adanya dukungan *NLP* yang matang, sebuah sistem kecerdasan buatan akan kesulitan menginterpretasikan maksud pengguna, terutama ketika dihadapkan pada variasi gaya bahasa masyarakat lokal yang dinamis [16].

2.4 Arsitektur *Transformers*

Sebelum meluasnya penggunaan arsitektur Transformers, model-model pemrosesan bahasa alami sebagian besar bertumpu pada jaringan saraf berulang seperti *Recurrent Neural Network (RNN)* dan *Long Short-Term Memory (LSTM)*. Meskipun cukup efektif untuk memproses data sekuensial, model berbasis RNN memiliki kelemahan mendasar dalam menangani dependensi jarak jauh (*long-range dependencies*) pada teks yang panjang. Selain itu, proses komputasi *RNN* yang bersifat sekuensial dari kiri ke kanan membuatnya tidak dapat dilatih secara paralel, sehingga membutuhkan waktu pelatihan yang sangat lama pada dataset skala besar [17].

Revolusi besar dalam bidang *NLP* terjadi ketika arsitektur Transformers diperkenalkan. Arsitektur ini sepenuhnya meninggalkan mekanisme jaringan saraf berulang dan menggantinya dengan mekanisme perhatian mandiri (*self-attention mechanism*). Mekanisme *self-attention* memungkinkan model untuk memeriksa seluruh kata dalam sebuah kalimat secara simultan dan menghitung bobot hubungan antar-kata tersebut tanpa mempedulikan jarak posisi mereka. Hal ini menghasilkan representasi konteks kata yang jauh lebih kaya dan akurat dibandingkan metode-metode pendahulunya [18].

Kelebihan fungsional lain dari arsitektur *Transformers* adalah kemampuannya untuk mendukung komputasi paralel secara penuh selama fase pelatihan model. Karena seluruh urutan kata diproses pada saat yang sama, proses komputasi dapat didistribusikan ke banyak *unit pemroses grafis (GPU)* sekaligus. Karakteristik ini membuka jalan bagi pengembangan model-model bahasa raksasa (*Large Language Models*) yang dilatih menggunakan korpus data teks berukuran gigabita hingga terabita, yang kemudian menjadi fondasi bagi teknologi *AI* generatif modern [19].

2.5 Model IndoBERT

Model *IndoBERT* merupakan varian mutakhir dari arsitektur *Bidirectional Encoder Representations from Transformers (BERT)* yang dirancang dan dilatih secara khusus untuk memahami bahasa Indonesia. *BERT* sendiri memiliki karakteristik unik berupa pelatihan dua arah (*bidirectional*), yang berarti model mempelajari konteks suatu kata berdasarkan kata-kata sebelum dan sesudahnya

secara bersamaan. Pendekatan kontekstual dua arah ini membuat model mampu membedakan makna kata yang sama namun berada dalam konteks kalimat yang berbeda (homonim) secara sangat presisi [20].

IndoBERT dikembangkan dengan melakukan prapelatihan (*pre-training*) menggunakan korpus teks bahasa Indonesia berskala besar, yang mencakup miliaran kata dari berbagai sumber seperti artikel berita, media sosial, Wikipedia bahasa Indonesia, hingga dokumen formal. Melalui proses prapelatihan yang intensif ini, *IndoBERT* berhasil menguasai karakteristik linguistik, sintaksis, semantik, serta struktur formal maupun informal dari bahasa Indonesia. Alhasil, model ini memiliki modal pengetahuan bahasa (*language understanding*) yang sangat kuat dibandingkan model bahasa universal lainnya [21].

Dalam siklus pengembangan aplikasi *AI* berbasis teks, *IndoBERT* bertindak sebagai *pre-trained* model yang siap pakai. Pengembang tidak perlu lagi membangun model bahasa dari nol yang memakan waktu dan biaya komputasi yang masif. Cukup dengan memanfaatkan pemahaman bahasa yang sudah dimiliki oleh *IndoBERT*, pengembang dapat melakukan proses penyesuaian parameter untuk tugas-tugas spesifik (*downstream tasks*), menjadikannya pilihan paling efisien dan akurat untuk membangun aplikasi cerdas berbahasa Indonesia [22].

2.6 Sistem Tanya Jawab (*Question Answering System*)

Sistem Tanya Jawab (*Question Answering System*) merupakan salah satu cabang terapan *NLP* tingkat lanjut yang berfokus pada kemampuan komputer

untuk memberikan jawaban yang relevan dan akurat terhadap pertanyaan yang diajukan oleh manusia dalam bahasa alami. Berbeda dengan mesin pencari (*search engine*) konvensional yang hanya memberikan daftar tautan situs web atau dokumen yang relevan, sistem tanya jawab langsung menyajikan informasi spesifik atau potongan teks yang dicari oleh pengguna, sehingga memangkas waktu pencarian secara drastis [23].

Secara umum, sistem tanya jawab berbasis *AI* dapat dikategorikan menjadi beberapa jenis pendekatan operasional. Pendekatan klasifikasi niat (*intent classification*) bekerja dengan cara mencocokkan pertanyaan pengguna ke dalam kategori-kategori niat tertentu yang telah didefinisikan sebelumnya, lalu mengeluarkan jawaban statis yang sesuai dengan niat tersebut. Sementara itu, pendekatan berbasis ekstraksi (*extractive QA*) bekerja secara lebih dinamis dengan cara memindai sebuah dokumen referensi teks, lalu mencari dan mengekstrak fragmen kalimat yang paling tepat untuk menjawab pertanyaan pengguna secara langsung [24].

Dalam implementasinya untuk layanan informasi daerah, pembangunan sistem tanya jawab membutuhkan pangkalan pengetahuan (*knowledge base*) yang terstruktur dengan baik. Kombinasi antara model bahasa canggih seperti *IndoBERT* dengan basis pengetahuan lokal yang komprehensif akan menghasilkan sistem informasi yang responsif. Keberhasilan dari sistem tanya jawab ini sangat dipengaruhi oleh kelengkapan variasi data latih yang diberikan serta kemampuan model dalam menangani ambiguitas kalimat interogatif dari pengguna [25].

2.7 *Fine-Tuning* dan Evaluasi Model *NLP*

Fine-tuning atau penalaan halus adalah sebuah proses dalam machine learning di mana sebuah model yang telah dilatih sebelumnya (*pre-trained model*) dilatih ulang menggunakan dataset baru yang berukuran lebih kecil dan spesifik untuk tugas tertentu. Dalam konteks penelitian ini, model *IndoBERT* yang sudah memiliki pengetahuan dasar bahasa Indonesia akan menjalani proses *fine-tuning* menggunakan dataset informasi spesifik Kabupaten Rokan Hulu. Proses ini bertujuan untuk menggeser bobot internal jaringan saraf pada model agar selaras dengan domain informasi yang baru tanpa menghilangkan kemampuan bahasa dasar yang telah dikuasai sebelumnya [26].

Selama fase *fine-tuning*, optimasi model dilakukan dengan menyesuaikan beberapa hiperparameter penting, seperti tingkat pembelajaran (*learning rate*), ukuran *batch* (*batch size*), dan jumlah iterasi pelatihan (*epochs*). Penentuan nilai hiperparameter yang tepat sangat krusial untuk mencegah terjadinya fenomena *overfitting*, di mana model menghafal dataset pelatihan terlalu detail sehingga gagal melakukan generalisasi dengan baik saat diuji menggunakan pertanyaan-pertanyaan baru yang belum pernah ditemuinya selama proses latihan [27].

Untuk mengukur performa dan efektivitas model *NLP* yang telah melalui proses *fine-tuning*, diperlukan tahapan evaluasi menggunakan metrik standar pengujian kuantitatif. Beberapa metrik yang paling umum digunakan adalah tingkat akurasi (*Accuracy*), *Precision* (kemampuan model memberikan jawaban benar dari keseluruhan tebakan benar), *Recall* (kemampuan model menemukan

kembali seluruh informasi benar), dan *F1-Score* yang merupakan nilai rata-rata harmonik antara *precision* dan *recall*. Nilai-nilai metrik inilah yang menjadi bukti ilmiah validitas performa bot cerdas yang dikembangkan [28].

2.8 Profil dan Kebutuhan Informasi Kabupaten Rokan Hulu

Kabupaten Rokan Hulu merupakan salah satu wilayah administratif strategis yang terletak di Provinsi Riau, Indonesia. Kabupaten yang dimekarkan pada tahun 1999 ini memiliki kekayaan budaya, nilai historis yang kuat, serta potensi ekonomi yang melimpah dari sektor perkebunan, pertanian, dan pariwisata religi maupun alam. Sebagai kawasan yang tengah berkembang, kebutuhan akan aksesibilitas data dan informasi yang cepat mengenai regulasi, destinasi wisata, serta struktur kelembagaan daerah di Rokan Hulu semakin meningkat seiring dengan laju modernisasi masyarakatnya [29].

Akses informasi yang merata di seluruh wilayah Kabupaten Rokan Hulu menjadi tantangan tersendiri bagi jajaran pemerintah daerah. Karakteristik geografis wilayah yang luas serta tersebar nya pusat-pusat pemukiman penduduk menuntut adanya pusat informasi digital yang inklusif dan dapat diakses dari mana saja melalui jaringan seluler. Kebutuhan utama masyarakat lokal maupun pendatang umumnya berkisar pada panduan prosedur pengurusan dokumen administrasi kependudukan, lokasi fasilitas kesehatan, jadwal kegiatan budaya, hingga pemetaan potensi investasi daerah [30].

Pengembangan *Intelligent Regional Bot (IRB)* yang dikhususkan untuk Kabupaten Rokan Hulu hadir sebagai jawaban ilmiah atas kebutuhan spasial

tersebut. Dengan mengintegrasikan pangkalan data karakteristik lokal Rokan Hulu ke dalam sistem kecerdasan buatan berbasis *IndoBERT*, jarak geografis dan keterbatasan jam operasional birokrasi konvensional dapat dipangkas. Penelitian ini menempatkan profil empiris Kabupaten Rokan Hulu bukan sekadar sebagai objek studi kasus, melainkan sebagai domain pangkalan data utama dalam mewujudkan digitalisasi daerah yang responsif [31].

2.9 Penelitian Terkait

Berikut adalah penelitian sebelumnya yang dijadikan acuan dan referensi dalam menyelesaikan masalah ini.

Tabel 2. 1 Tabel Penelitian Sebelumnya

No	Nama	Tahun	Judul	Hasil
1	Budi Santoso, Rina Wati, M. Fikri [32]	2026	Implementasi <i>Fine-Tuning IndoBERT</i> pada <i>Asisten Virtual</i> Pariwisata Daerah	Berdasarkan hasil penelitian, penerapan <i>fine-tuning</i> pada model IndoBERT berhasil meningkatkan performa asisten virtual dalam memahami pertanyaan terkait destinasi wisata lokal. Model mencapai nilai <i>F1-score</i> sebesar 96,2% dalam klasifikasi niat (<i>intent classification</i>) pengguna. Pendekatan ini terbukti jauh melampaui sistem <i>rule-based</i> konvensional, terutama dalam menangani struktur kalimat yang tidak baku dan variasi kata (sinonim). Peneliti

				menyarankan pengembangan selanjutnya agar menambahkan fitur pengenalan bahasa daerah untuk meningkatkan kealamian interaksi..
2.	Ahmad Rinaldi, Siti Aisyah, Lukman Hakim [33].	2026	Evaluasi Kinerja <i>Chatbot E-Government</i> Menggunakan <i>Natural Language Processing</i> Berbasis <i>Transformers</i>	Penelitian ini mengevaluasi arsitektur berbasis <i>Transformers</i> untuk layanan kependudukan tingkat kota. Hasil uji menunjukkan bahwa model bahasa memiliki tingkat akurasi 94,8% dengan waktu inferensi rata-rata 0,3 detik per pesan. Kemampuan model dalam memahami konteks dua arah (<i>bidirectional</i>) membuat <i>chatbot</i> mampu merespons pertanyaan ambigu dengan tepat. Kendati demikian, model masih rentan terhadap <i>overfitting</i> pada kelas data yang minoritas, sehingga direkomendasikan penggunaan teknik augmentasi teks pada penelitian berikutnya.
3.	Dian Puspita, Rizky Ramadhan, Andi Yusuf [34]	2025	Klasifikasi Niat Pengguna pada Aplikasi Layanan Masyarakat Menggunakan <i>IndoBERT</i>	Hasil pengujian membuktikan bahwa <i>IndoBERT</i> sangat efektif untuk memetakan keluhan dan pertanyaan masyarakat ke dalam kategori layanan publik yang tepat (seperti kesehatan, infrastruktur, dan administrasi). Dengan

				akurasi mencapai 95,5%, sistem ini mampu mengurangi waktu tunggu respons layanan secara signifikan. Namun, kesalahan klasifikasi masih kerap terjadi jika pengguna menggabungkan dua topik keluhan dalam satu kalimat panjang. Disarankan untuk memecah kalimat majemuk pada tahap prapengolahan data (<i>preprocessing</i>).
4.	Kevin Pratama, Nanda Safitri [35]	2025	Optimasi Ekstraksi Informasi Layanan Publik Melalui Pendekatan <i>Question Answering</i> Berbasis <i>Deep Learning</i>	Penelitian ini berhasil merancang sistem <i>Question Answering</i> (QA) ekstraktif untuk menggali informasi langsung dari dokumen PDF peraturan daerah. Model mampu menemukan rentang jawaban (<i>answer span</i>) yang akurat di dalam teks panjang dengan metrik <i>Exact Match</i> (EM) sebesar 82% dan <i>F1-score</i> 88%. Meskipun menjanjikan, sistem ini membutuhkan daya komputasi server yang cukup besar saat memproses dokumen yang melebihi 50 halaman.
5.	Sarah Meliana, Taufik Hidayat, Vina Anggraini [36]	2025	Pemanfaatan <i>Transfer Learning</i> <i>IndoBERT</i> untuk Deteksi Kebutuhan Informasi Kependuduka	Penggunaan teknik <i>transfer learning</i> pada <i>pre-trained model IndoBERT</i> terbukti mempercepat proses konvergensi saat pelatihan model <i>chatbot</i> desa. Model dapat dilatih hanya dalam waktu 45 menit

			n	menggunakan GPU spesifikasi menengah, namun menghasilkan akurasi hingga 93%. Penelitian ini menegaskan bahwa penggunaan <i>IndoBERT</i> sangat ramah sumber daya untuk skala penelitian kecil. Disarankan untuk memperluas <i>dataset</i> mencakup istilah-istilah sosiokultural setempat.
6.	Yoga Saputra, Desi Ratnasari, M. Ihsan [37]	2024	Komparasi Model <i>BERT</i> dan <i>TF-IDF</i> dalam Pembuatan Bot Layanan Akademik Interaktif	Hasil komparasi menunjukkan keunggulan absolut dari arsitektur BERT dibandingkan metode ekstraksi fitur tradisional TF-IDF. Sementara TF-IDF hanya mencapai akurasi 76% karena gagal menangkap makna semantik, varian BERT mencatatkan akurasi 91%. BERT terbukti mampu membedakan makna kata yang sama dalam konteks berbeda (homonim). Keterbatasan penelitian terletak pada jumlah korpus data latih yang masih terbatas pada satu fakultas saja.
7.	Fahmi Idris, Rika Amelia [38]	2024	Rancang Bangun <i>Chatbot</i> Interaktif untuk Promosi Kebudayaan Lokal	Sistem <i>chatbot</i> yang dibangun menggunakan NLP berhasil diterima dengan baik oleh pengguna dengan tingkat kepuasan <i>User Acceptance Test (UAT)</i> sebesar 89%. Bot dinilai

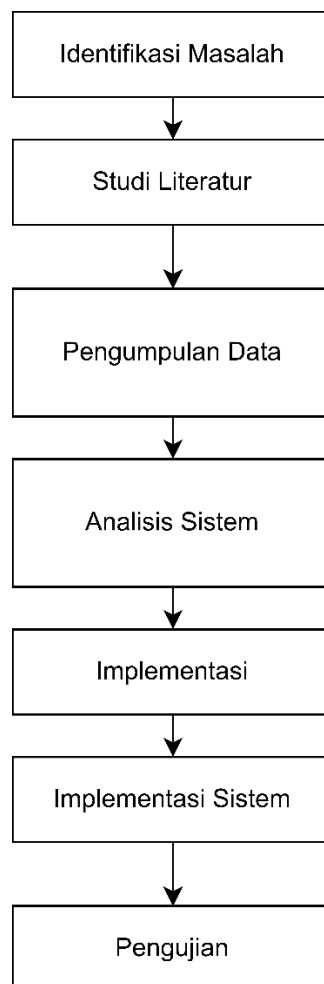
			Menggunakan Algoritma <i>NLP</i>	interaktif dan mampu menjawab sejarah kebudayaan lokal dengan narasi yang luwes. Evaluasi teknis menunjukkan model masih kesulitan mendeteksi kesalahan ketik (<i>typo</i>) ekstrem, sehingga penelitian selanjutnya disarankan untuk mengintegrasikan algoritma koreksi ejaan (<i>spell checker</i>) sebelum teks diproses oleh model <i>NLP</i> .
8.	Ratna Dwi, Arya Wiguna, Bima Sakti [39]	2024	Implementasi <i>Natural Language Processing</i> pada Pengarsipan Keluhan Publik di Tingkat Kabupaten	Analisis pemrosesan teks menggunakan <i>NLP</i> terbukti mampu mengekstraksi kata kunci penting dari paragraf keluhan warga secara otomatis. Model ini mendapatkan tingkat presisi sebesar 88% dalam mengategorikan urgensi pesan. Keberhasilan ini membantu instansi mengotomatisasi pendataan, meskipun model belum mampu memahami sentimen sarkasme yang sering dilontarkan pengguna di media sosial.
9.	M. Zidan, Ayu Permatasari [40]	2023	Tinjauan Efektivitas <i>Chatbot</i> Pemerintahan Berbasis <i>Rule-Based</i> dan <i>Machine Learning</i>	Berdasarkan analisis komparatif, <i>chatbot</i> berbasis <i>machine learning</i> (<i>NLP</i>) menunjukkan fleksibilitas operasional yang jauh lebih superior dibandingkan <i>rule-based</i> .

				Pendekatan <i>rule-based</i> mengalami <i>fallback rate</i> (kegagalan menjawab) hingga 40% saat dihadapkan pada bahasa gaul, sedangkan <i>NLP</i> mampu menekan angka kegagalan tersebut hingga di bawah 12%. Riset ini merekomendasikan migrasi penuh dari sistem aturan kaku menuju sistem berbasis <i>machine learning</i> untuk <i>e-government</i>.
10.	Indah Pertiwi, Surya Dharma, Rudianto [41]	2023	Implementasi <i>Question Answering System</i> pada Portal Berita Daerah Menggunakan Arsitektur Jaringan Saraf Tiruan	Penelitian merancang agen tanya-jawab untuk menjawab pertanyaan berbasis artikel berita daerah secara instan. Sistem mencapai akurasi 79% dalam mengekstraksi jawaban benar. Tantangan utama yang ditemukan adalah panjangnya paragraf berita yang menyebabkan model kehilangan fokus konteks. Solusi yang disarankan untuk riset ke depan adalah mengimplementasikan mekanisme <i>attention</i> yang lebih mutakhir seperti yang terdapat pada keluarga arsitektur Transformers..

BAB 3

METODOLOGI PENELITIAN

Penelitian ini dilaksanakan dengan mengikuti tahapan demi tahapan yang saling berhubungan. Setiap tahapan dijelaskan dalam metode penelitian yang disusun secara sistematis, terstruktur, dan ditampilkan dalam bentuk skema yang jelas. Berikut tahapan-tahapan penelitian dapat dilihat pada Gambar 3.1 dibawah:



Gambar 3. 1 Alur Metodologi Penelitian

Penjelasan dari tahapan-tahapan penelitian pada Gambar 3.1 dapat dilihat pada penjelasan dibawah ini:

3.1 Identifikasi Masalah

Masyarakat maupun wisatawan di Kabupaten Rokan Hulu masih menghadapi kesulitan dalam mengakses informasi daerah secara terpadu, cepat, dan akurat, khususnya terkait topik layanan publik, pariwisata, dan profil wilayah. Selama ini, penyampaian informasi daerah cenderung masih tersebar di berbagai website yang statis atau sangat bergantung pada pelayanan manual yang dibatasi oleh jam operasional instansi. Selain itu, belum tersedia sistem pelayanan informasi berbasis teknologi cerdas yang mampu memberikan jawaban otomatis dan kontekstual secara real-time sesuai dengan ragam pertanyaan masyarakat. Penggunaan chatbot konvensional yang bersifat rule-based (berbasis aturan) juga memiliki keterbatasan yang signifikan dalam memahami keragaman struktur kalimat dan variasi bahasa sehari-hari pengguna. Di sisi lain, pemanfaatan teknologi pemrosesan bahasa alami (*Natural Language Processing*) termutakhir seperti model *IndoBERT* belum banyak diimplementasikan sebagai solusi pusat informasi cerdas berskala regional di Indonesia. Hal ini menjadi tantangan tersendiri dalam merancang sistem *Intelligent Regional Bot (IRB)* yang tidak hanya mampu menyediakan pangkalan data informasi daerah secara terstruktur, tetapi juga cerdas dalam memahami niat (*intent*) dan merespons pertanyaan pengguna secara natural dan akurat.

3.2 Studi Literatur

Dalam penelitian ini, akan dilakukan peninjauan terhadap beberapa jurnal, artikel ilmiah, dan skripsi terdahulu yang relevan. Selain itu, dilakukan juga pengumpulan data melalui studi dokumen pada portal resmi pemerintah daerah, publikasi instansi, serta literatur terkait profil dan informasi publik Kabupaten Rokan Hulu. Langkah ini bertujuan untuk memperkuat landasan teori serta mempelajari secara mendalam konsep-konsep di balik teknologi *Natural Language Processing (NLP)*, pemodelan bahasa menggunakan arsitektur *IndoBERT*, serta mekanisme perancangan *Intelligent Regional Bot (IRB)* sebagai sistem layanan informasi daerah.

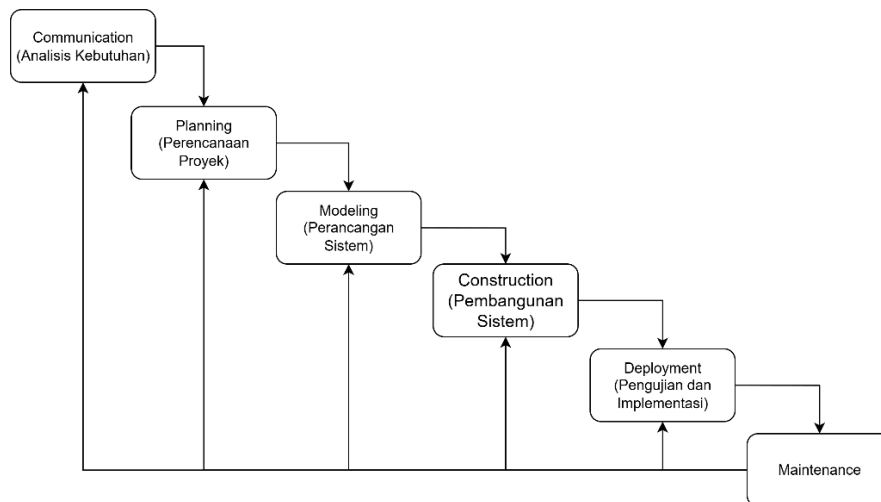
3.3 Pengumpulan Data

Pada tahap ini, peneliti akan mengumpulkan dan mempersiapkan data yang diperoleh dari situs web resmi pemerintah daerah, dokumen publik, artikel, jurnal, serta berbagai literatur terkait mengenai “informasi Kabupaten Rokan Hulu”, yang meliputi profil wilayah, potensi pariwisata, dan prosedur layanan publik. Data yang telah terkumpul tersebut kemudian akan diproses dan disusun sedemikian rupa agar dapat digunakan sebagai dataset terstruktur. Dataset ini nantinya akan menjadi pangkalan pengetahuan (*knowledge base*) sekaligus rujukan jawaban utama bagi sistem *Intelligent Regional Bot (IRB)* berbasis model *IndoBERT* yang akan dikembangkan oleh peneliti.

3.4 Analisis Sistem

Analisis sistem merupakan suatu proses penjabaran dari sistem penyebaran informasi yang berjalan saat ini dengan maksud untuk mengidentifikasi berbagai permasalahan (seperti akses data yang statis dan terbatas) serta mengevaluasi kebutuhan pengguna. Tujuannya adalah untuk merancang dan membangun sebuah sistem baru, yaitu *Intelligent Regional Bot (IRB)*, yang dapat mengatasi kendala pada sistem sebelumnya. Tahapan ini mencakup analisis alur kerja program (*flowchart*), analisis kebutuhan spesifikasi sistem, analisis masukan (*input* berupa teks pertanyaan dari pengguna), serta analisis keluaran (*output* berupa respons informasi cerdas dari bot) agar hasil pengembangan sesuai dengan fungsionalitas yang diharapkan.

Dalam pengembangan sistem *IRB* ini, pendekatan rekayasa perangkat lunak yang diterapkan adalah model *Waterfall*. Model *Waterfall* merupakan metode yang menerapkan pendekatan pengembangan perangkat lunak secara sistematis dan sekuensial (berurutan). Siklus hidup pengembangan pada metode ini bergerak maju secara bertahap melalui fase analisis kebutuhan (*analysis*), perancangan arsitektur dan antarmuka (*design*), penulisan kode program dan implementasi model *IndoBERT* (*coding*), pengujian performa fungsionalitas dan akurasi *NLP* (*testing*), hingga pemeliharaan sistem (*maintenance*).



Gambar 3. 2 Metode *Waterfall*

Gambar 3.2 menjelaskan metode ini terdiri dari beberapa tahapan yang disusun secara berurutan dan sistematis, dimulai dari analisis kebutuhan hingga pemeliharaan sistem.

1. *Communication* (Analisis Kebutuhan)

mana dilakukan identifikasi kebutuhan sistem dengan mengumpulkan data dan dokumen terkait informasi Kabupaten Rokan Hulu, seperti potensi pariwisata, profil daerah, dan panduan layanan publik. Kumpulan informasi ini akan diproses menjadi dataset yang menjadi dasar pengetahuan bagi chatbot cerdas.

2. *Planning* (Perencanaan Proyek)

Mencakup pemilihan teknologi dan perangkat lunak yang akan digunakan, seperti bahasa pemrograman *Python*, *framework* untuk pembuatan antarmuka (misalnya *Flask*, *FastAPI*, atau *Streamlit*), serta penyiapan arsitektur model *IndoBERT* sebagai otak pemrosesan bahasa alami (*NLP*).

Pada tahap ini juga disusun jadwal pengerjaan dan tahapan pengembangan sistem.

3. Modeling (Perancangan Sistem)

Meliputi perancangan alur kerja sistem (flowchart), desain antarmuka pengguna (*User Interface*), serta pemodelan basis data (dataset). Pada tahap ini juga dirancang skenario prapengolahan teks (*text preprocessing*) dan arsitektur *fine-tuning* pada model *IndoBERT* agar sistem mampu mengklasifikasikan pertanyaan atau niat (*intent*) pengguna dengan tepat.

4. Construction (Pembangunan Sistem)

Yaitu tahap penulisan kode program (coding) dan implementasi sistem secara langsung. Dataset informasi Rokan Hulu dilatihkan ke dalam model *IndoBERT* melalui proses *fine-tuning*. Setelah model dilatih dan dievaluasi secara internal, model *NLP* tersebut diintegrasikan dengan antarmuka chatbot agar dapat menerima masukan teks dari pengguna dan menampilkan jawaban yang sesuai dengan konteks.

5. Deployment (Pengujian dan Implementasi)

Sistem diuji menggunakan berbagai variasi pertanyaan dari pengguna untuk memastikan bot dapat memahami konteks bahasa dan memberikan jawaban yang relevan serta akurat. Evaluasi dilakukan menggunakan metrik pengujian algoritma *NLP* (seperti *Accuracy*, *Precision*, *Recall*, atau *F1-Score*) serta pengujian fungsionalitas antarmuka oleh calon pengguna.

6. *Maintenance* (Pemeliharaan Sistem)

Mencakup pembaruan dataset jika terdapat perubahan informasi terbaru mengenai Kabupaten Rokan Hulu ,misalnya perubahan jam layanan publik atau penambahan lokasi wisata .pelatihan ulang (*retraining*) model *IndoBERT* dengan data baru, serta perbaikan bug pada aplikasi berdasarkan umpan balik (*feedback*) dari masyarakat pengguna. Tahap ini bertujuan agar *chatbot* tetap responsif dan informasinya selalu mutakhir.

3.5 Implementasi

Implementasi merupakan tahapan dalam siklus pengembangan sistem di mana rancangan arsitektur dan solusi algoritma yang telah dibuat mulai diterjemahkan ke dalam bentuk kode program agar dapat beroperasi secara fungsional pada lingkungan nyata. Ini melibatkan penerapan ide, strategi, dan pemodelan kecerdasan buatan ke dalam sistem operasional yang siap digunakan. Pada tahap ini, peneliti menggunakan platform berbasis web sebagai wadah implementasi utama untuk aplikasi bot cerdas.

1. Pengembangan *Intelligent Regional Bot (IRB)*

Sistem *Intelligent Regional Bot (IRB)* ini dikembangkan menggunakan bahasa pemrograman *Python* sebagai inti dari pemrosesan logika sistem (backend). Proses pemahaman bahasa alami, ekstraksi fitur teks, dan penentuan jawaban sepenuhnya mengandalkan model *IndoBERT* yang telah dilatih secara khusus (*fine-tuning*) menggunakan dataset informasi Kabupaten Rokan Hulu. Untuk memberikan pengalaman interaksi yang

intuitif bagi masyarakat, antarmuka pengguna (*frontend*) dibangun menggunakan kombinasi teknologi web yaitu *HTML*, *CSS*, dan *JavaScript*. Antarmuka aplikasi ini dirancang secara efisien dan terpusat untuk tingkat pengguna umum (*User*), di mana fokus utama sistem adalah menyajikan jawaban informatif secara langsung tanpa memerlukan fitur pencatatan riwayat analisis (*history analysis*) interaksi sebelumnya.

3.6 Implementasi Sistem

Beberapa komponen pendukung yang berperan penting dalam implementasi sistem diantaranya adalah perangkat keras (*hardware*) dan perangkat lunak (*software*). Adapun spesifikasi dari perangkat keras dan perangkat lunak yang digunakan sebagai berikut:

1. Perangkat keras (*hardware*), antara lain:

Processor : *Intel(R) Core (TM) i3-4030U CPU @ 1.90GHz (4 CPUs), ~1.9GHz*

Memory : 4.00 GB

System type : *Windows 10 Pro 64-bit (10.0, build 17763)*

Harddisk : 500 GB

2. Perangkat lunak (*software*), antara lain:

Sistem operasi : *Windows 10*

Tools : *Google Colab*

3.7 Pengujian

Tahap pengujian ini bertujuan untuk mengukur tingkat kelayakan dan performa dari sistem Intelligent Regional Bot (IRB) yang telah dikembangkan. Pengujian difokuskan pada evaluasi tingkat kecocokan antara respons yang dihasilkan oleh bot dengan data referensi aktual terkait informasi Kabupaten Rokan Hulu. Proses komputasi pada pengujian ini akan dieksekusi menggunakan platform Google Colab, di mana model IndoBERT dijalankan dan dievaluasi secara otomatis. Kualitas sistem diukur dengan membandingkan keluaran model terhadap jawaban referensi (ground truth) menggunakan metrik evaluasi algoritma pemrosesan bahasa alami, seperti akurasi kecocokan teks atau BERTScore. Nilai evaluasi yang diperoleh dari tahapan ini akan menjadi indikator utama untuk menilai sejauh mana bot cerdas tersebut mampu memahami konteks pertanyaan pengguna dan mereproduksi informasi daerah secara presisi.