

BAB II

LANDASAN TEORI

2.1. Tinjauan Pustaka

2.1.1. Natural Language Processing (NLP)

Natural Language Processing (NLP) merupakan cabang dari kecerdasan buatan yang memungkinkan komputer untuk memahami, menginterpretasikan, dan menghasilkan bahasa manusia secara alami. Dalam konteks pengelolaan dokumen digital, NLP digunakan untuk meningkatkan kemampuan sistem dalam memahami konteks dan makna dari permintaan pengguna, sehingga dapat memberikan respons yang lebih relevan dan akurat (Puspitasari et al., 2024)

2.1.1.1. Definisi dan Komponen NLP

NLP terdiri dari beberapa komponen utama yang bekerja secara berurutan untuk memproses bahasa alami:

1. **Tokenisasi:** Memecah teks menjadi unit-unit kecil seperti kata atau frasa.
2. **Stemming dan Lemmatization:** Mengubah kata ke bentuk dasarnya untuk analisis lebih lanjut.
3. **Part-of-Speech Tagging:** Menentukan kelas kata (kata benda, kata kerja, dll.) dari setiap token.

Komponen-komponen ini memungkinkan sistem untuk memahami konteks dan makna dari input pengguna, yang penting dalam pengembangan chatbot yang responsif dan akurat.

2.1.1.2. Fungsi dan Penerapan NLP

Penerapan NLP mencakup berbagai bidang seperti pengklasifikasian teks, analisis sentimen, summarization, dan pengenalan entitas. Salah satu penerapan utamanya adalah dalam pembuatan chatbot yang mampu memahami permintaan pengguna melalui teks alami, kemudian memberikan tanggapan berdasarkan konteks dokumen yang tersimpan dalam sistem.

Menurut (Salamun et al., 2024) NLP memungkinkan sistem informasi untuk melakukan pencarian dokumen tidak hanya berbasis kata kunci literal, tetapi juga berdasarkan makna dan niat pengguna. Ini sangat penting dalam pengelolaan arsip digital yang jumlahnya besar dan bervariasi, terutama di instansi pemerintahan.

2.1.1.3. Tantangan dalam Implementasi NLP

Meskipun sudah banyak dikembangkan, NLP masih menghadapi tantangan, seperti:

1. Keragaman Bahasa: Bahasa Indonesia memiliki banyak bentuk morfologis dan idiomatik.
2. Format Dokumen Tidak Konsisten: Dokumen yang diolah seringkali tidak terstruktur dengan baik.

3. Keterbatasan Berbahasa Lokal: Banyak model NLP yang masih dominan dalam bahasa Inggris.

(Khatib Sulaiman et al., 2024) menekankan perlunya pengembangan NLP yang berbasis pada konteks lokal agar dapat memberikan hasil yang lebih akurat, terutama dalam pelayanan publik dan administrasi pemerintahan.

2.1.2. Chatbot

Chatbot berbasis NLP dirancang untuk memahami dan merespons pertanyaan pengguna dalam bahasa alami. Teknologi ini memungkinkan chatbot untuk melakukan analisis teks, memahami intent pengguna, dan memberikan jawaban yang sesuai berdasarkan data yang tersedia. (Fachrizal, 2024)

2.1.3. Pengelolaan Dokumen Digital

Pengelolaan dokumen digital melibatkan proses penyimpanan, pengorganisasian, dan pencarian dokumen dalam format digital. Integrasi NLP dalam sistem pengelolaan dokumen digital memungkinkan pencarian berbasis konteks dan makna, bukan hanya berdasarkan kata kunci. Hal ini meningkatkan akurasi dan efisiensi dalam menemukan dokumen yang relevan. (Rahmadani et al., 2024)

2.1.4. Pemodelan Sistem UML

Unified Modeling Language (UML) adalah bahasa pemodelan standar yang digunakan untuk menspesifikasikan, memvisualisasikan, membangun, dan mendokumentasikan artefak dari sistem perangkat lunak. UML menyediakan berbagai

jenis diagram yang membantu dalam proses perancangan sistem, seperti Use Case Diagram, Activity Diagram, Sequence Diagram, dan Class Diagram.

Menurut (Wayahdi et al., 2023). Diagram UML yang digunakan mencakup Use Case Diagram untuk menggambarkan interaksi antara aktor dan sistem, Activity Diagram untuk menunjukkan alur aktivitas, Sequence Diagram untuk memodelkan interaksi antar objek berdasarkan waktu, dan Class Diagram untuk mendeskripsikan struktur kelas dalam sistem.

2.1.4.1. Usecase

Use Case Diagram adalah salah satu jenis diagram dalam Unified Modeling Language (UML) yang digunakan untuk menggambarkan fungsionalitas sistem dari perspektif pengguna (aktor). Diagram ini berfungsi untuk menjelaskan bagaimana pengguna (aktor) berinteraksi dengan sistem serta menggambarkan layanan yang diberikan sistem kepada aktor.

Menurut(Wayahdi et al., 2023) terdapat beberapa komponen utama dalam sebuah Use Case Diagram, yaitu:

1. Aktor (Actor)

Aktor adalah entitas eksternal yang berinteraksi langsung dengan sistem. Aktor bisa berupa pengguna manusia, sistem lain, atau perangkat eksternal yang berperan dalam proses bisnis sistem.

2. Use Case

Use case menggambarkan fungsi atau layanan yang diberikan sistem kepada aktor. Contohnya dalam sistem chatbot adalah "Mencari Dokumen", "Mengunggah Berkas", dan "Menerima Jawaban dari Chatbot".

3. Asosiasi (Association)

Merupakan hubungan antara aktor dengan use case. Digambarkan dengan garis lurus yang menghubungkan aktor ke use case yang digunakannya.

4. Sistem Boundary (Batas Sistem)

Batas sistem adalah kotak yang membungkus semua use case, menunjukkan mana proses yang merupakan bagian dari sistem dan mana yang berada di luar sistem.

5. Relasi Include dan Extend

Include digunakan ketika suatu use case selalu melibatkan use case lain sebagai bagian dari prosesnya.

Extend digunakan untuk menunjukkan bahwa suatu use case dapat diperluas dengan perilaku tambahan pada kondisi tertentu.

2.1.4.2. Diagram Activity

Diagram activity merupakan salah satu jenis diagram dalam Unified Modeling Language (UML) yang digunakan untuk merepresentasikan alur kerja atau aktivitas

dalam sistem secara terstruktur. Diagram ini memodelkan urutan aktivitas yang dilakukan oleh sistem atau pengguna dari awal hingga akhir.

Menurut (Wayahdi et al., 2023) activity diagram sangat berguna untuk menggambarkan logika proses sistem, terutama untuk sistem yang memiliki alur aktivitas yang kompleks seperti dalam penerapan chatbot berbasis NLP yang menangani interaksi dinamis dengan pengguna.

Adapun komponen-komponen utama dari activity diagram adalah:

1. Initial Node (Titik Awal)

Digambarkan dengan lingkaran berisi titik hitam, sebagai tanda dimulainya suatu alur proses.

2. Action/Activity (Tindakan/Aktivitas)

Menunjukkan pekerjaan atau proses yang dilakukan, biasanya digambarkan dengan persegi panjang bersudut tumpul.

3. Decision Node (Keputusan)

Menandakan titik percabangan alur berdasarkan kondisi tertentu, digambarkan dengan belah ketupat.

4. Merge Node (Penggabungan)

Digunakan untuk menggabungkan beberapa alur menjadi satu kembali.

5. Fork dan Join Node (Pecahan dan Penggabungan Paralel)

Fork membagi alur menjadi beberapa jalur paralel, sedangkan Join menyatukannya kembali.

6. Final Node (Titik Akhir)

Digambarkan sebagai lingkaran dengan lingkaran kecil di dalamnya, yang menunjukkan akhir dari proses.

2.1.5. Software dan pemrograman terkait

2.1.5.1 Visual Studio Code

Menurut (Mutiara et al., 2024) Visual Studio Code digunakan sebagai teks editor utama dalam pengembangan sistem ini karena kemampuannya mendukung berbagai bahasa pemrograman, ringan, dan memiliki fitur-fitur tambahan melalui ekstensi yang mempercepat proses pengembangan.

2.1.5.2 Navicat

Menurut (Fathulloh et al., 2021) Navicat adalah aplikasi besutan PremiumSoft CyberTech Ltd yang berfungsi manajemen database. Pengguna aplikasi ini bisa menggunakan berbagai jenis database seperti (MySQL, Oracle, PostgreSQL, SQLite, dll) dan bisa diimplementasikan di beragam operating system seperti Linux, Windows, dan Macintosh. User Interface pada Navicat mengadopsi hierarki database dari phpMyAdmin dan memuat beberapa kustomisasi agar mempermudah pengguna dalam mengonfigurasi database-nya. Navicat merupakan software yang premium, jadi 14 sebelum Anda menggunakan aplikasi ini, Anda diharuskan membeli lisensi resminya

dari Navicat. Dan pada instalasinya, akan diminta konfirmasi pembelian dengan memasukkan kode lisensi saat melakukan instalasi software tersebut, Sebagai database interface Navicat sangat memudahkan penggunaanya dalam mengeksekusi Data Definition Language dan dalam mengeksekusi Data Manipulation Language dari berbagai jenis Database Management System (DBMS). Dengan menggunakan aplikasi ini, Anda tidak diharuskan menghafal semua syntax dari query karena Navicat akan mengeksekusi semua query yang diperlukan hanya dengan melalui klik pada aplikasinya

2.1.5.3 PHP

PHP merupakan script untuk pemrograman script web server-side, script yang membuat dokumen HTML secara on the fly, maksudnya dokumen HTML yang dihasilkan dari suatu aplikasi bukan dokumen HTML yang dibuat dengan menggunakan editor teks atau editor HTML.(Sains et al., 2022)

2.1.5.4 Javascript

Javascript adalah kerangka kerja yang siap digunakan dan sistematis pada Javascript dengan bantuan sekumpulan library, agar dapat mempermudah para pengembang dalam membuat situs web berbasis JavaScript, seperti membuat fungsi, prosedur, serta fitur lainnya memakai bantuan library yang telah ada sehingga tidak harus untuk menuliskan kembali sumber kode-nya. Framework Javascript memiliki 20 dua tipe, yaitu: Framework Javascript Frontend dan Framework Javascript Backend. Framework Javascript Frontend menghadirkan dukungan pada pengembangan

tampilan antarmuka situs web yang interaktif, menarik agar mempermudah pengguna dalam mengakses situs web serta meningkatkan pengalaman user.(Christopher et al., 2022)

2.1.5.5 Database dan DBMS

Menurut(Riyan Dirgantara et al., 2023) DBMS adalah perangkat lunak yang digunakan untuk mengelola basis data. DBMS bertanggung jawab atas penyimpanan, pengambilan, pembaruan, dan penghapusan data dalam basis data. DBMS menyediakan antarmuka dan alat untuk mengatur struktur basis data, memanipulasi data, dan menjaga keamanan dan integritas data. Manajemen basis data melibatkan berbagai teknologi dan metode, termasuk database management systems (DBMS), database design, data modeling, data mining, dan data warehousing. Sistem manajemen basis data modern telah berkembang pesat selama beberapa dekade terakhir, dan 19 sekarang tersedia berbagai jenis DBMS yang dapat disesuaikan dengan kebutuhan bisnis yang berbeda.

2.1.5.6 MY SQL

MySQL adalah DBMS yang open source dengan dua bentuk lisensi, yaitu Free Software (perangkat lunak bebas) dan Shareware (perangkat lunak berpemilik yang penggunaannya terbatas). Jadi MySQL adalah database server yang gratis dengan lisensi GNU General Public License (GPL) sehingga dapat Anda pakai untuk keperluan pribadi atau komersial tanpa harus membayar lisensi yang ada.(Kalsum Siregar et al., 2024)

2.1.6. Fast Api

Dalam pengembangan sistem chatbot berbasis Natural Language Processing (NLP), dibutuhkan sebuah framework yang mampu menangani permintaan pengguna secara cepat, efisien, dan fleksibel. Untuk itu, penulis menggunakan FastAPI, sebuah web framework modern berbasis Python yang dirancang khusus untuk membangun aplikasi web berbasis REST API.

FastAPI dipilih karena kemampuannya dalam menangani permintaan secara asynchronous, mendukung validasi data secara otomatis melalui Pydantic, serta mudah diintegrasikan dengan model machine learning seperti BERT. Dalam sistem ini, FastAPI berperan sebagai backend utama yang mengelola input pengguna, memproses prediksi intent menggunakan model NLP, serta menjalankan logika pencarian dan ringkasan dokumen digital.

2.1.6.1 Cara Kerja FastAPI

A. User Interaction (Input Bahasa Alami)

1. Pengguna mengirim pertanyaan atau perintah menggunakan bahasa alami (contoh: *“Tolong carikan berkas atas nama Wisnu”*).
2. Input ini dikirim melalui tampilan frontend (biasanya berupa formulir, chatbot UI, atau REST client seperti Postman).

B. FastAPI Endpoint Terima Permintaan

1. FastAPI memiliki endpoint POST /nlp yang menerima data berupa question melalui body JSON.
2. Permintaan ini diproses oleh kelas BaseModel (QueryModel) untuk validasi input.

C. Proses Intent Classification (NLP dengan BERT)

1. Input user dikirim ke model NLP (misalnya indobenchmark/indobert-base-p1) yang telah dilatih sebelumnya.
2. Model memprediksi intent dari input tersebut, seperti: greeting, search_by_criteria, read_pdf_content, summarize_pdf, dll.

D. Eksekusi Aksi Berdasarkan Intent

Setelah intent dikenali, server menjalankan aksi sesuai intent:

1. Jika `search_by_criteria`, maka akan mencari nama pada database permohonan.
2. Jika `read_pdf_content`, maka akan membaca isi PDF menggunakan pustaka PyMuPDF (fitz).
3. Jika `summarize_pdf`, maka isi PDF diringkas menggunakan model summarization (facebook/bart-large-cnn).

E. Respon Balik ke User

1. Setelah selesai chatbot akan menampilkan output berdasarkan intent dari user tersebut

2.2. Kajian Penelitian Terdahulu

Penelitian terkait dan yang relevan dengan topik “Penerapan Metode Natural Language Processing pada Pengelolaan Berkas Digital menggunakan Chatbot di Pusdatin Dinas PPKUKM” Berikut beberapa paper penelitiannya:

2.2.1. Paper 1

Penelitian yang di lakukan oleh (Kambala, 2023) Berjudul "Natural Language Processing for IT Documentation and Knowledge Management" dipublikasikan pada tanggal 25 Februari 2023 dalam jurnal International Journal of Scientific Research and Management (IJSRM)

2.2.1.1. Tujuan Penelitian

bertujuan untuk mengeksplorasi bagaimana teknologi Natural Language Processing (NLP) dapat digunakan untuk meningkatkan dokumentasi dan manajemen pengetahuan di bidang teknologi informasi (TI). Fokus utama dari penelitian ini adalah menganalisis peran NLP dalam meningkatkan proses pencarian, klasifikasi, dan pembuatan dokumentasi IT secara otomatis. Selain itu, penelitian ini juga mengevaluasi berbagai teknik NLP, seperti Named Entity Recognition (NER), Topic Modeling, dan pencarian berbasis Transformer, untuk melihat seberapa efektif teknologi tersebut dalam konteks manajemen pengetahuan. Studi ini turut menyertakan studi kasus nyata dari perusahaan multinasional yang menggunakan solusi NLP untuk mengatasi tantangan dalam pengelolaan dokumen teknis.

2.2.1.2. Metodologi yang Digunakan

Metodologi yang digunakan dalam penelitian ini meliputi tinjauan literatur terhadap berbagai pendekatan NLP terkini, serta studi kasus yang menggambarkan implementasi nyata NLP di lingkungan perusahaan IT. Dalam studi kasus tersebut, peneliti menggunakan beberapa teknik NLP seperti Named Entity Recognition untuk mengekstraksi istilah teknis dari teks tidak terstruktur, Semantic Search menggunakan model SBERT untuk meningkatkan akurasi pencarian berdasarkan makna, Topic Modeling untuk pengelompokan dokumen berdasarkan topik tertentu

2.2.1.3. Temuan Utama

Temuan utama dari penelitian ini menunjukkan bahwa penggunaan NLP secara signifikan meningkatkan efisiensi dokumentasi dan layanan IT. Contohnya, waktu rata-rata pencarian dokumen berkurang dari 12 menit menjadi hanya 3 menit, sedangkan waktu penyelesaian tiket layanan IT menurun drastis dari 48 jam menjadi 12 jam.

2.2.1.4. Kesimpulan Penelitian

Sebagai kesimpulan, penelitian ini menyatakan bahwa NLP berpotensi merevolusi dokumentasi dan manajemen pengetahuan IT dengan cara yang lebih otomatis, efisien, dan mudah diakses. Meskipun demikian, masih ada beberapa tantangan dalam penerapan teknologi ini, seperti kualitas data yang rendah, kebutuhan adaptasi pada istilah teknis IT, keterbatasan integrasi dengan sistem lama, dan masalah keamanan data.

2.2.2. Paper 2

Penelitian yang dilakukan oleh (Khatib Sulaiman et al., 2024) Berjudul "Penggunaan Teknologi Natural Language Processing dalam Sistem Chatbot untuk Peningkatan Layanan Informasi Administrasi Publik" yang dipublikasikan pada Indonesian Journal of Computer Science, Vol. 13 No. 1 Tahun 2024 telah terakreditasi Sinta 4

2.2.2.1. Tujuan Penelitian

Penelitian ini bertujuan untuk meningkatkan kualitas layanan informasi administrasi publik di Pemerintahan Desa Tosari dengan membangun sistem chatbot berbasis teknologi *Natural Language Processing* (NLP). Chatbot yang dikembangkan diintegrasikan dengan aplikasi Telegram untuk memberikan layanan informasi administrasi secara cepat, efisien, dan mudah diakses oleh masyarakat.

2.2.2.2. Metodologi yang Digunakan

Metode yang digunakan adalah pendekatan waterfall, yang meliputi tahapan analisis kebutuhan, perancangan sistem, implementasi, pengujian, dan pemeliharaan.

2.2.2.3. Temuan Utama

Temuan utama dari penelitian ini menunjukkan bahwa chatbot yang dikembangkan mampu memberikan respons yang akurat dan sesuai dengan kebutuhan informasi administrasi masyarakat. Nilai rata-rata SUS mencapai 86, yang menempatkannya pada kategori "Excellent".

2.2.2.4. Kesimpulan

Penelitian ini menyimpulkan bahwa chatbot berbasis NLP yang dikembangkan menggunakan DialogFlow dan diintegrasikan dengan Telegram dapat meningkatkan efisiensi dan kualitas layanan informasi administrasi publik di desa. Sistem ini dinilai sangat baik oleh pengguna berdasarkan hasil pengujian usability dan pengalaman pengguna.

2.2.3. Paper 3

Pada Penelitian yang dilakukan oleh (Verolina et al., 2024) Berjudul “Implementasi Chatbot untuk Membantu Wisatawan dalam Mencari Wisata yang Berada di Kabupaten Semarang dengan Menggunakan Metode Natural Language Processing (NLP)” yang dipublikasikan pada jurnal Riset Sistem Informasi dan Teknik Informatika (JURASIK) pada tahun 2024 telah terakreditasi Sinta 4

2.2.3.1. Tujuan Penelitian

Penelitian ini bertujuan untuk mengembangkan sebuah chatbot berbasis teknologi *Natural Language Processing (NLP)* yang dapat membantu wisatawan dalam mencari informasi tentang destinasi wisata yang berada di Kabupaten Semarang. Dengan semakin meningkatnya kunjungan wisata pascapandemi dan pentingnya sektor pariwisata dalam pertumbuhan ekonomi lokal, chatbot ini diharapkan dapat memberikan kemudahan akses informasi bagi wisatawan kapan pun dan di mana pun mereka berada

2.2.3.2. Metodologi yang Digunakan

Metodologi penelitian yang digunakan Algoritma NLP digunakan untuk memahami pertanyaan pengguna melalui proses tokenisasi, stemming, dan lemmatization, kemudian chatbot memberikan jawaban berdasarkan database destinasi wisata yang telah disiapkan.

2.2.3.3. Temuan utama

Temuan utama dari penelitian menunjukkan bahwa chatbot berhasil memberikan jawaban yang sesuai dan akurat untuk semua skenario pengujian yang diuji, Pengguna dapat mengakses chatbot melalui Telegram, dan interaksi percakapan yang dilakukan mampu menanggapi pertanyaan secara kontekstual, mulai dari sapaan pembuka, penjawaban terkait destinasi dan fasilitas, hingga menangani input yang tidak valid melalui pesan fallback.

2.2.3.4. Kesimpulan

Kesimpulan dari penelitian ini menyatakan bahwa implementasi chatbot menggunakan metode NLP mampu memberikan solusi informasi yang cepat dan interaktif bagi wisatawan. Chatbot ini terbukti efektif dalam membantu pengunjung memperoleh informasi mengenai wisata Kabupaten Semarang secara otomatis, akurat, dan tanpa batasan waktu.