

BAB I PENDAHULUAN

1.1 Latar Belakang

PT Tabel Data Informatika merupakan perusahaan pengembang perangkat lunak yang bertempat di Bandung, Jawa Barat. Sebagai perusahaan yang bergerak dibidang teknologi dengan menyediakan layanan IT Service, IT Consultant, dan Talent Development berbasis website dan mobile app.

PT Tabel Data Informatika memiliki proyek internal untuk mengembangkan aplikasi website untuk keperluan *Continuous Auditing and Continuous Monitoring (CACM)*. *Continuous Auditing and Continuous Monitoring (CACM)* adalah cara bagaimana pengguna untuk melihat dan menggunakan data yang berpotensi menciptakan keputusan yang menguntungkan dalam perspektif bisnis [1].

Saat ini, aplikasi website tersebut hanya menampilkan berupa *chart* sebagai visualisasi data untuk keperluan audit dan pemantauan berkelanjutan. Aplikasi ini belum menyediakan kebebasan bagi pengguna untuk melihat data lain diluar visualisasi yang telah ditentukan, sehingga membatasi kemampuan eksplorasi data secara mendalam.

Eksplorasi data pada sistem saat ini hanya bisa dilakukan di dalam *database* yang memerlukan pengetahuan mengenai bahasa khusus, yaitu *Structured Query Language (SQL)*. Oleh karena itu, pengguna dari divisi non-teknis harus mengajukan permintaan kepada tim IT saat ingin melakukan eksplorasi data secara lebih mendalam terhadap data di luar visualisasi yang telah disediakan pada sistem. Hal ini tidak hanya memperlambat proses pengambilan keputusan, tetapi juga menambah beban kerja bagi tim IT karena harus menangani permintaan eksplorasi data secara manual.

Untuk mengatasi masalah tersebut dibutuhkan sebuah solusi yang dapat menjembatani keterbatasan pengguna non-teknis dalam mendapatkan informasi dari dalam *database* secara mandiri tanpa harus memahami sintaks SQL. Salah satu solusi yang dapat diterapkan adalah menambahkan fitur chatbot pada aplikasi website tersebut, yang memungkinkan pengguna bertanya dalam bahasa alami dan mendapatkan jawaban langsung dari *database*.

Chatbot ini bekerja dengan bantuan *Natural Language Processing (NLP)*. NLP adalah bidang ilmu komputer dan kecerdasan buatan yang berfokus pada interaksi antara komputer dan bahasa manusia [2].

Dalam perkembangannya, NLP saat ini banyak didukung oleh teknologi *Large Language Model (LLM)*. LLM sendiri adalah model kecerdasan buatan yang mampu

memahami serta menghasilkan bahasa alami. Model LLM umumnya dilatih menggunakan data teks dalam jumlah besar dan menggunakan teknik pembelajaran mendalam atau *deep learning* untuk memahami struktur dan pola bahasa [3].

Dalam konteks pengelolaan *database*, LLM yang dapat dimanfaatkan untuk membangun sistem tanya jawab berbasis data, dikenal dengan istilah *Text-to-SQL*. Salah satu model LLM yang dirancang khusus untuk melakukan kebutuhan ini adalah DefogLlama, yang memungkinkan integrasi teknologi NLP ke dalam aplikasi seperti *chatbot*.

Untuk menjaga kerahasiaan data dan informasi perusahaan, proyek akhir ini tidak menggunakan data asli milik PT Tabel Data Informatika. Sebagai gantinya, digunakan data *dummy* berupa *database* toko media digital atau yang dikenal dengan DBChinook. DBChinook merupakan contoh *database* sumber terbuka yang dapat digunakan dalam pembelajaran atau pengujian sistem berbasis *database* sebagai pengganti *database* asli perusahaan, yang mendukung *database* seperti MySQL, Oracle, dan PostgreSQL.

Dalam proyek ini, DBChinook disimulasikan sebagai *database* yang merepresentasikan sistem penjualan produk digital seperti musik dan album. *Database* ini mencakup berbagai tabel seperti transaksi penjualan, data pelanggan, dan genre musik. Struktur data pada DBChinook digunakan sebagai dasar simulasi untuk pengembangan *chatbot* yang dibangun dalam proyek ini.

1.2 Rumusan Masalah dan Solusi

Berdasarkan latar belakang di atas, maka rumusan masalah untuk proyek akhir ini adalah pengguna non-teknis harus mengajukan permintaan kepada tim IT setiap kali membutuhkan data dari *database* yang dapat memperlambat proses pengambilan keputusan dan meningkatkan beban kerja tim IT.

Berdasarkan rumusan masalah tersebut, maka solusi yang dapat diberikan adalah pengembangan *chatbot* yang dapat membantu pengguna mendapatkan data dari *database* hanya dengan mengajukan pertanyaan.

1.3 Tujuan

Berdasarkan rumusan masalah dan solusi diatas, proyek akhir ini memiliki tujuan membangun *chatbot* berbasis web yang mampu memberikan data dari *database* berdasarkan pertanyaan pengguna menggunakan Large Language Model (LLM).

1.4 Batasan Masalah

Beberapa hal yang menjadi batasan masalah dalam proyek akhir ini diantaranya adalah:

1. *Chatbot* hanya mampu menjawab pertanyaan yang relevan dengan tabel dan kolom yang terintegrasi.
2. *Chatbot* hanya mampu menangani satu konteks pertanyaan, tidak bisa menangani berdasarkan pertanyaan sebelumnya.
3. Bahasa yang digunakan dalam pertanyaan adalah bahasa Indonesia dan bahasa Inggris.
4. *Chatbot* hanya mampu menjawab pertanyaan dari pengguna untuk menampilkan data dan tidak bisa memanipulasi data seperti *insert*, *update*, dan *delete* untuk menghindari risiko perubahan data yang tidak disengaja.
5. Proyek akhir ini tidak mencakup pengembangan model baru, melainkan hanya mengimplementasikan dan menguji model yang sudah ada.
6. *Chatbot* hanya mendukung *database* Chinook, sehingga tidak dapat digunakan untuk *database* lain tanpa penyesuaian tambahan.

1.5 Penjadwalan Kerja

Rangkaian kerja yang dilakukan pada proyek ini dijelaskan secara terstruktur berdasarkan tahapan-tahapan yang telah dilakukan setiap bulannya. Tabel berikut menggambarkan rincian kegiatan yang dilakukan selama pengerjaan proyek ini:

Tabel 1. 1 Tabel penjadwalan kerja

No	Deskripsi Kerja	November 2024				Desember 2024				Januari 2025				Februari 2025			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Analisis kebutuhan	█	█	█	█												
2	Persiapan data				█												
3	Pemilihan model LLM				█												
4	Pengembangan prototype sistem					█	█	█	█	█	█	█	█	█	█	█	█
5	Pengujian dan evaluasi prototype					█	█	█	█	█	█	█	█	█	█	█	█

No	Deskripsi Kerja	November 2024				Desember 2024				Januari 2025				Februari 2025			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
6	Implementasi sistem final																