

1. Pendahuluan

Latar Belakang

Penggunaan obat herbal sebagai alternatif pengobatan terus meningkat secara global, didorong oleh keyakinan akan keamanannya, harga terjangkau, dan ketersediaan bahan alami [1][2]. Obat herbal dianggap sebagai pengobatan awal sebelum menggunakan obat konvensional, terutama dalam mengobati penyakit ringan atau sedang [3]. Popularitas ini juga berpotensi mendukung *Sustainable Development Goals* (SDGs) 2030, melalui solusi pengobatan yang mudah diakses [4].

Namun, tantangan utama seiring popularitas ini adalah kurangnya akses terhadap informasi valid mengenai khasiat, dosis, efek samping, dan informasi lainnya mengenai produk herbal, yang merupakan bagian dari pengobatan herbal. Pengguna sering kesulitan mendapatkan jawaban akurat dan cepat atas pertanyaan spesifik mengenai produk herbal. Informasi yang tersebar di berbagai sumber seringkali tidak terverifikasi. Tantangan ini menunjukkan pentingnya sistem informasi yang baik dalam pengobatan tradisional agar mampu menyediakan informasi yang akurat bagi pengguna [5].

Sistem Tanya Jawab (*Question Answering System* - QAS) menawarkan solusi dengan kemampuannya memahami pertanyaan berbahasa alami dan memberikan jawaban relevan [6]. Untuk domain terstruktur seperti informasi obat herbal, pendekatan berbasis ontologi menjadi pilihan tepat [7]. Ontologi merepresentasikan pengetahuan secara terstruktur melalui konsep dan relasinya, memungkinkan QAS memberikan informasi yang akurat dan relevan.

Kendala utama QAS berbasis ontologi adalah pemahaman pertanyaan pengguna dan juga kompleksitas dalam pencarian informasi dari ontologi [8]. Untuk mengatasi kendala tersebut, penelitian ini menggunakan *Named Entity Recognition* (NER) yang efektif dalam mengidentifikasi entitas-entitas penting dalam sebuah pertanyaan [9]. NER berperan untuk mengekstraksi kata kunci seperti nama produk, penyakit, atau komposisi yang kemudian akan digunakan untuk pencarian informasi dari ontologi.

Selain itu, *Support Vector Machine* (SVM) digunakan karena kinerjanya yang handal dalam klasifikasi teks pertanyaan [10]. SVM digunakan dengan tujuan agar sistem dapat memahami pertanyaan pengguna dan mencari jawaban yang spesifik berdasarkan pengetahuan dalam ontologi yang tersedia. Pada penelitian ini digunakan berbagai representasi teks, yaitu TF-IDF dan FastText dengan melalui beberapa uji coba untuk meningkatkan optimalisasi kinerja SVM.

Large Language Models (LLMs) hadir menawarkan keuntungan yang besar untuk meningkatkan kemampuan QAS, melalui pemahaman konteks yang lebih baik dan hasil jawaban yang lebih natural [11]. Informasi ontologi dapat digunakan sebagai konteks untuk meningkatkan pengetahuan LLM [12]. Dengan integrasi ini, informasi berdasarkan fakta dari ontologi dapat dirangkum dan dijadikan informasi yang lebih natural. Kombinasi antara informasi dari ontologi produk obat herbal dan LLM diharapkan dapat memberi jawaban yang akurat dan relevan bagi pengguna.

Di era digital, akses informasi semakin mudah melalui berbagai platform pesan. Sistem tanya jawab ini akan diimplementasikan melalui telegram bot untuk mempermudah akses pengguna terhadap informasi valid produk obat herbal.

Topik dan Batasannya

Penelitian ini berfokus pada mengembangkan Sistem Tanya Jawab (QAS) berbasis ontologi produk obat herbal dengan memanfaatkan *Named Entity Recognition* (NER) sebagai model ekstraksi kata kunci dan mengoptimalkan kinerja *Support Vector Machine* (SVM) sebagai klasifikasi pertanyaan untuk pencarian informasi pada ontologi, yang selanjutnya diimplementasikan pada *Large Language Model* (LLM).

Pengembangan sistem ini memiliki beberapa batasan. Pertama, cakupan pengetahuan ontologi dibatasi pada informasi produk obat herbal dari website BPOM, meliputi nama obat, manfaat, komposisi, kemasan, dan produsen, sehingga tidak mencakup berbagai aspek obat herbal secara umum. Kedua, mayoritas (80%) data pertanyaan untuk melatih klasifikasi SVM dibuat secara manual karena cakupan pengetahuan ontologi yang ada dan keterbatasan data berbahasa Indonesia mengenai produk herbal. Batasan-batasan ini merupakan penyederhanaan yang diperlukan dengan sumber daya dan waktu yang tersedia.

Tujuan

Penelitian ini bertujuan untuk merancang dan mengimplementasikan sebuah arsitektur sistem tanya jawab yang spesifik mengintegrasikan *Support Vector Machine* (SVM) dan *Named Entity Recognition* (NER) untuk pencarian informasi dari ontologi produk obat herbal. Informasi tersebut kemudian diimplementasikan sebagai konteks bagi *Large Language Model* (LLM).

Tujuan selanjutnya yaitu mengoptimalkan kinerja klasifikasi SVM. Pada penelitian ini akan diuji beberapa vektorisasi teks dan pengujian penggunaan pra-pemrosesan teks. Vektorisasi teks yang digunakan yaitu TF-IDF, FastText PreTrained Model, FastText Manual Trained, dan IndoBERT menggunakan data pertanyaan. Ketiga vektorisasi tersebut diuji dengan menggunakan pra-pemrosesan teks dan tidak menggunakan pra-pemrosesan teks untuk meningkatkan optimalisasi kinerja model klasifikasi SVM.

Selain itu, penelitian ini bertujuan mewujudkan sistem tanya jawab pada telegram bot untuk meningkatkan aksesibilitas informasi produk obat herbal kepada pengguna. Keberhasilan penelitian ini akan diukur melalui evaluasi pada NER dan klasifikasi SVM dengan beberapa vektorisasi teks, guna menghasilkan sistem yang dapat memberikan jawaban yang akurat dan relevan dari ontologi produk obat herbal.

Organisasi Tulisan

Struktur penulisan ini terbagi menjadi lima bagian. Bagian kedua menyajikan penelitian yang dilakukan sebelumnya, beserta penjelasan konsep-konsep dasar agar lebih memahami keseluruhan penelitian ini. Bagian ketiga menjelaskan secara rinci seluruh pembangunan sistem tanya jawab dan penjelasan alur pengguna dalam mendapatkan informasi produk obat herbal. Bagian keempat berfokus pada evaluasi dari klasifikasi SVM dan NER melalui berbagai matriks untuk dianalisis efektivitasnya. Bagian terakhir yaitu kesimpulan mengenai keseluruhan penelitian yang telah dilakukan.