

## 1. Pendahuluan

### Latar Belakang

Sejak Desember 2019, seluruh dunia telah digemparkan dengan munculnya wabah Covid-19 tidak terkecuali di Indonesia. *World Health Organization* (WHO) melaporkan peningkatan jumlah kasus *coronavirus(2019-nCov)-infected pneumonia* (NCIP) atau Covid-19 yang pertama kali diidentifikasi di Kota Wuhan, Provinsi Hubei, China dengan 4 kasus pertama yang dilaporkan pada 29 Desember 2019 [1]. Pada 2 Maret 2020, Kementerian Kesehatan Republik Indonesia telah melaporkan 2 kasus terkonfirmasi Covid-19 dan meningkat menjadi 1.285 kasus pada 29 Maret 2020. Munculnya wabah Covid-19 menjadi tantangan utama bagi masyarakat Indonesia agar selalu mendapatkan informasi yang terbaru dan paling relevan mengenai perkembangan covid-19 di Indonesia [2].

Upaya untuk mengumpulkan informasi tentang perkembangan Covid-19, sangat erat kaitannya dengan sistem pencari (*search engine*). Hal ini juga terlihat pada beberapa situs web yang telah membuat halaman khusus untuk pertanyaan umum atau biasa disebut *frequently asked questions* (FAQ). Namun, masyarakat mungkin kesulitan menemukan jawaban pada *search engine* ataupun halaman FAQ tersebut dikarenakan hasil dari *search engine* masih berupa kumpulan dokumen dan belum bisa menyediakan informasi jawaban kepada pengguna secara langsung. Oleh karena itu, keberadaan suatu sistem informasi yang dikenal dengan *Question Answering System* (QAS) yang dapat menerima inputan pengguna berupa kalimat pertanyaan dan menghasilkan sebuah jawaban secara cepat, ringkas, serta relevan sangat diperlukan.

*Recognizing Question Entailment* (RQE) adalah salah satu metode yang dapat diimplementasikan pada *Question Answering System*. RQE adalah komponen penting dalam pemahaman bahasa alami dan dalam menjawab pertanyaan. Menurut Harabagiu dan Hickle, RQE dapat memungkinkan sistem untuk mengidentifikasi jawaban yang benar dengan akurasi yang lebih besar dibanding dengan menggunakan *keyword* atau menggunakan *pattern based methods* [3].

Oleh karena itu dalam tugas akhir ini akan diimplementasikan *Question Answering System* berbahasa Indonesia dengan metode *Recognizing Question Entailment* menggunakan *dataset* berupa pertanyaan masyarakat yang berasal dari akun-akun Twitter yang kredibel tentang Covid-19 seperti dokter, satuan tugas Covid-19, dan Kementerian Kesehatan.

### Topik dan Batasannya

Sistem Tanya Jawab atau *Question Answering System* (QAS) merupakan sebuah sistem yang dapat menerima pertanyaan yang berisi informasi yang diperlukan untuk menghasilkan jawaban yang diinginkan. QAS dirancang untuk dapat memenuhi kebutuhan informasi manusia dengan cepat dan tepat seperti situasi ketika berbicara dengan asisten virtual, interaksi dengan mesin pencari, dan melakukan *query* pada *database* [4]. Pada kasus tersebut, peneliti membangun Sistem Tanya Jawab dengan pendekatan *Recognizing Question Entailment* untuk menentukan apakah makna dari suatu teks pertanyaan dapat disimpulkan dari teks pertanyaan lainnya [5].

Pada penelitian ini mengambil referensi *dataset* pertanyaan yang dikumpulkan dari bulan Maret 2020 hingga Mei 2021 berupa pertanyaan masyarakat dalam Bahasa Indonesia yang berasal dari akun-akun Twitter yang kredibel tentang covid-19 seperti dokter, satuan tugas covid-19, dan Kementerian Kesehatan.

### Tujuan

Penelitian tugas akhir ini bertujuan untuk mengimplementasikan metode *Recognizing Question Entailment* (RQE) pada *Question Answering System* Covid-19 Bahasa Indonesia menggunakan beberapa algoritma *supervised learning* yang biasa digunakan untuk RQE, diantaranya adalah *Support Vector Machine* (SVM), *Logistic Regression*, *Naïve Bayes*, dan J48.

Tujuan selanjutnya dari tugas akhir ini adalah untuk menganalisis hasil implementasi dari metode *Recognizing Question Entailment* pada *Question Answering System* Covid-19 Bahasa Indonesia.

### Organisasi Tulisan

Setelah bagian pendahuluan, jurnal tugas akhir ini dilanjutkan dengan bagian studi terkait yang di dalamnya menjelaskan terkait penelitian-penelitian sebelumnya pada domain kesehatan. Pada bagian kedua jurnal ini dijelaskan mengenai hal-hal yang masih berkaitan terhadap batasan pekerjaan yang dilakukan. Pada bagian ketiga jurnal ini akan dijelaskan mengenai beberapa metode yang digunakan dan sistem yang dibangun seperti metode *Recognizing Question Entailment*, algoritma *supervised learning*, dan juga tahap pra-proses secara lebih mendalam. Pada bagian keempat jurnal ini akan menjelaskan hasil evaluasi dari beberapa skenario pengujian serta

analisis hasil pengujian dengan beberapa metrik evaluasi seperti *precision*, *recall*, dan *f-measure*. Pada bagian kelima jurnal ini akan dijelaskan mengenai kesimpulan dari penelitian ini dan saran untuk penelitian selanjutnya.