

BAB I PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi pengenalan suara dan pemrosesan bahasa alami (*NLP*) telah membawa dampak signifikan dalam berbagai bidang, termasuk pendidikan, kesehatan, dan layanan pelanggan. Model-model kecerdasan buatan (*AI*) terbaru mampu mengubah input suara menjadi teks dengan akurasi tinggi, memungkinkan interaksi yang lebih alami antara manusia dan mesin.

Salah satu model terbaru dalam pengenalan suara adalah *Whisper* yang dikembangkan oleh OpenAI. *Whisper* adalah sistem pengenalan ucapan otomatis yang dilatih menggunakan 680.000 jam data multibahasa dan multitugas yang dikumpulkan dari *web*. Model ini mampu melakukan transkripsi multibahasa, identifikasi bahasa, dan terjemahan ucapan ke bahasa Inggris [1].

Di sisi lain, dalam domain pemrosesan bahasa alami, model *Sentence Bidirectional Encoder Representations from Transformers (SBERT)* telah diperkenalkan untuk menghasilkan representasi kalimat yang bermakna secara semantik. *SBERT* memodifikasi jaringan BERT pra-terlatih menjadi arsitektur *Siamese* dan *triplet*, memungkinkan perhitungan kesamaan semantik antar kalimat dengan efisien [2].

Al-Qur'an, sebagai kitab suci umat Islam, terdiri dari 114 surah dan 6.236 ayat. Banyak umat Muslim yang ingin mencari atau mengingat ayat tertentu berdasarkan potongan teks atau bunyi yang diingat. Namun, proses pencarian ayat yang spesifik dapat menjadi tantangan, terutama jika hanya mengandalkan ingatan parsial atau tanpa mengetahui nomor surah dan ayat.

Implementasi teknologi pengenalan suara dan pencocokan teks semantik dapat mempermudah proses ini. Dengan sistem yang mampu mentranskripsi ucapan pengguna dan mencocokkannya dengan ayat Al-Qur'an yang relevan, aksesibilitas dan interaksi dengan teks suci ini dapat ditingkatkan.

Dalam kehidupan sehari-hari, banyak umat Muslim menghadapi kesulitan dalam mencari ayat Al-Qur'an yang mereka ingat hanya sebagian. Misalnya, para penghafal Al-Qur'an (*hafizh*) yang sedang mengikuti kompetisi atau mengulang hafalan seringkali lupa lokasi ayat secara spesifik. Hal serupa juga dialami oleh

penceramah, guru agama, atau akademisi yang membutuhkan referensi ayat dalam kajian tafsir dan hukum Islam, tetapi hanya mengingat potongan frasanya tanpa mengetahui nomor surah dan ayat. Selain itu, masyarakat umum yang ingin mencari ayat terkait suatu topik, seperti kesabaran, kejujuran, atau larangan riba, sering mengalami kesulitan karena keterbatasan dalam mengingat teks lengkapnya.

Di era digital, kebutuhan akan pencarian ayat berbasis suara juga semakin meningkat, terutama dengan maraknya dakwah melalui media sosial seperti YouTube, TikTok, dan Instagram. Banyak pengguna yang mendengar kutipan ayat dari video ceramah atau konten edukasi Islam tanpa mengetahui sumbernya secara pasti. Pencarian ayat Al-Qur'an saat ini masih mengandalkan metode berbasis teks, yang sering kali tidak efektif jika pengguna hanya mengingat sebagian kata atau memiliki kesalahan dalam penulisan. Oleh karena itu, diperlukan sebuah sistem yang dapat mengonversi suara menjadi teks secara akurat serta mencocokkannya dengan ayat yang relevan, sehingga pencarian ayat Al-Qur'an menjadi lebih cepat dan efisien bagi berbagai kalangan.

Whisper adalah model pengenalan ucapan otomatis yang dirancang untuk mentranskripsi ucapan multibahasa dengan akurasi tinggi. Model ini menggunakan arsitektur *encoder-decoder Transformer*, di mana input audio diubah menjadi *spektrogram log-Mel* sebelum diproses oleh *encoder*. *Decoder* kemudian memprediksi teks yang sesuai, termasuk tugas-tugas seperti identifikasi bahasa dan penandaan waktu frasa [1].

Sentence Bidirectional Encoder Representations from Transformers (SBERT) adalah modifikasi dari model *BERT* yang dirancang untuk menghasilkan embedding kalimat yang bermakna secara semantik, di mana *BERT* sendiri merupakan arsitektur berbasis *Transformer* yang menggunakan mekanisme *self-attention* untuk memahami hubungan antar kata dalam suatu konteks. Dengan menggunakan arsitektur jaringan *Siamese*, *SBERT* dapat menghitung kesamaan antara dua kalimat dengan efisien, yang sangat berguna dalam tugas-tugas seperti pencarian semantik dan penemuan parafrasa. [2].

Beberapa penelitian telah dilakukan dalam domain pengenalan ucapan dan pencocokan teks semantik. Namun, penerapan spesifik teknologi ini untuk pencarian ayat Al-Qur'an berbasis suara masih terbatas. Sebagai contoh, meskipun

Whisper telah menunjukkan kemampuan transkripsi yang kuat, beberapa penelitian menunjukkan bahwa model ini dapat menghasilkan "halusinasi" atau teks yang tidak akurat dalam kondisi tertentu [3].

Selain itu, meskipun *SBERT* telah berhasil dalam berbagai aplikasi pencocokan teks, penerapannya dalam konteks teks keagamaan seperti Al-Qur'an memerlukan penelitian lebih lanjut untuk memastikan akurasi dan relevansi.

Penelitian ini bertujuan untuk mengembangkan sistem yang mengintegrasikan *Whisper* dan *SBERT* untuk memungkinkan pengguna mencari dan mencocokkan ayat Al-Qur'an melalui input suara. Dengan mentranskripsi ucapan pengguna menggunakan *Whisper* dan mencocokkan teks hasil transkripsi dengan ayat Al-Qur'an menggunakan *SBERT*, diharapkan sistem ini dapat memberikan hasil yang akurat dan relevan, sehingga mempermudah interaksi dan pemahaman pengguna terhadap Al-Qur'an.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, permasalahan yang akan dibahas dalam penelitian ini adalah sebagai berikut:

1. Bagaimana cara mengimplementasikan model *Whisper* untuk mengenali suara pengguna dan mengubahnya menjadi teks dengan tingkat akurasi yang tinggi?
2. Bagaimana cara menerapkan model *SBERT* untuk mencocokkan teks hasil transkripsi dengan ayat Al-Qur'an yang paling relevan?
3. Seberapa akurat sistem dalam mengenali suara pengguna dan menemukan ayat yang sesuai berdasarkan transkripsi suara?

1.3 Batasan Masalah

Agar penelitian lebih terfokus dan terarah, terdapat beberapa batasan dalam penelitian ini, yaitu:

1. Model pengenalan suara yang digunakan adalah *Whisper*, tanpa membandingkannya dengan model lain.
2. Model pencocokan teks yang digunakan adalah *SBERT*, dengan perbandingan terhadap metode pencocokan berbasis kata kunci secara sederhana.

3. *Dataset* ayat Al-Qur'an yang digunakan berasal dari teks terjemahan resmi dan teks Arab Al-Qur'an yang telah tersedia dalam bentuk digital.
4. Penelitian ini tidak membahas aspek tafsir atau makna mendalam dari ayat-ayat yang dicocokkan, tetapi hanya fokus pada pencocokan semantik berdasarkan teks.
5. Evaluasi dilakukan berdasarkan tingkat akurasi transkripsi suara dan pencocokan ayat yang sesuai.

1.4 Tujuan

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengembangkan sistem yang dapat mengenali suara pengguna dan mengubahnya menjadi teks menggunakan model *Whisper*.
2. Mengimplementasikan model *SBERT* untuk mencocokkan teks hasil transkripsi dengan ayat Al-Qur'an yang paling relevan.
3. Mengevaluasi performa sistem berdasarkan tingkat akurasi transkripsi suara dan keberhasilan pencocokan ayat Al-Qur'an.

1.5 Manfaat

Hasil dari penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Manfaat Akademik:

- a) Menambah referensi dan kontribusi dalam bidang *NLP* dan pengenalan suara, khususnya dalam konteks pencarian teks berbasis suara.
- b) Dapat dijadikan acuan bagi penelitian selanjutnya dalam pengembangan sistem pencocokan ayat berbasis *AI*.

2. Manfaat Praktis:

- a) Mempermudah umat Muslim dalam mencari ayat Al-Qur'an hanya dengan mengucapkan potongan ayat atau frasa yang mereka ingat.
- b) Dapat digunakan dalam aplikasi edukasi Al-Qur'an berbasis teknologi, seperti aplikasi penghafalan dan pencarian ayat.
- c) Berpotensi untuk diterapkan dalam sistem pencarian berbasis suara di masjid, lembaga pendidikan Islam, atau aplikasi dakwah digital.