

BAB I

PENDAHULUAN

1.1. Latar Belakang Penelitian

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. 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 seringkali 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.

Perkembangan dakwah di era digital telah membuka peluang baru dalam penyebaran nilai-nilai Islam melalui berbagai platform media sosial. Sebagaimana diberitakan oleh *BeritaSatu* (2025) [1], dakwah di media sosial dipandang sebagai salah satu implementasi dari QS. Al-Hijr ayat 94, yaitu perintah untuk menyampaikan ajaran Islam secara terbuka. Media sosial memungkinkan pesan dakwah menjangkau audiens yang lebih luas, lintas wilayah, dan lintas generasi. Hal ini menegaskan bahwa transformasi digital bukan hanya berdampak pada sektor sosial-ekonomi, tetapi juga pada cara umat Islam mengakses dan memahami ajaran agamanya. Dengan demikian, integrasi teknologi dalam penyampaian dakwah menjadi kebutuhan yang tidak dapat dihindari.

Sejalan dengan hal tersebut, laporan *Digital 2025* yang dirilis oleh *We Are Social* dan *Meltwater* memperlihatkan dominasi aplikasi berbasis mobile seperti

YouTube, TikTok, WhatsApp, Instagram, dan Facebook sebagai platform dengan jumlah pengguna aktif bulanan dan waktu penggunaan tertinggi sepanjang periode September–November 2024. Tren ini menunjukkan bahwa masyarakat global semakin bergantung pada media digital berbasis audio-visual untuk mencari informasi maupun hiburan. Kondisi ini memperkuat urgensi penelitian mengenai sistem pencarian ayat Al-Qur'an berbasis suara, yang sejalan dengan pola konsumsi digital masyarakat modern. Kehadiran sistem ini diharapkan dapat menjadi sarana inovatif dalam mendukung dakwah digital, sekaligus memudahkan umat Islam dalam menemukan dan memahami ayat Al-Qur'an melalui input suara yang praktis dan relevan dengan kebutuhan masa kini [2].

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.

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 artifisial (KA) 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. *Whisper* juga merupakan 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 [3].

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. *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 [4].

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 [5]. 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.

Untuk mendukung proses transkripsi dan pencocokan semantik ayat Al-Qur'an, penelitian ini menggunakan dua jenis *Dataset*, yaitu *Dataset* suara pengguna dan teks ayat Al-Qur'an. *Dataset* suara dikumpulkan secara langsung dari

partisipan yang membacakan ayat-ayat Al-Qur'an dalam format audio sebagai *input* model *Whisper*. Sementara itu, *Dataset* teks ayat Al-Qur'an diperoleh dari *platform Kaggle*, dengan judul "Al-Quran (*Arabic and Indonesian Translation*)" yang disusun oleh ronnieaban. *Dataset* ini menyediakan teks ayat dalam bahasa Arab dan terjemahannya dalam Bahasa Indonesia, yang digunakan sebagai referensi dalam proses pencocokan semantik menggunakan model SBERT. Pemanfaatan kedua *Dataset* ini memungkinkan skenario pengujian yang realistis dan komprehensif terhadap performa sistem dalam mencocokkan ayat Al-Qur'an hanya berdasarkan *input* suara.

Dari latar belakang tersebut, maka penulis tertarik melakukan penelitian mengenai "**Implementasi *Whisper* Dan *Sentence Bidirectional Encoder Representations From Transformers* Untuk Pencocokan Ayat Al-Qur'an Dari Suara**". Dimana penelitian ini bertujuan untuk mengembangkan model dengan mengintegrasikan *Whisper* dan SBERT yang 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 model ini dapat memberikan hasil yang akurat dan relevan, sehingga mempermudah interaksi dan pemahaman pengguna terhadap Al-Qur'an.

1.2. Perumusan Masalah Penelitian

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

1. Bagaimana cara mengimplementasikan model *Whisper* untuk mengubah suara pengguna 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 secara semantik?
3. Bagaimana pengukuran model *Whisper* dan SBERT dalam mendukung proses pencocokan ayat Al-Qur'an dari *input* suara?

1.3. Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

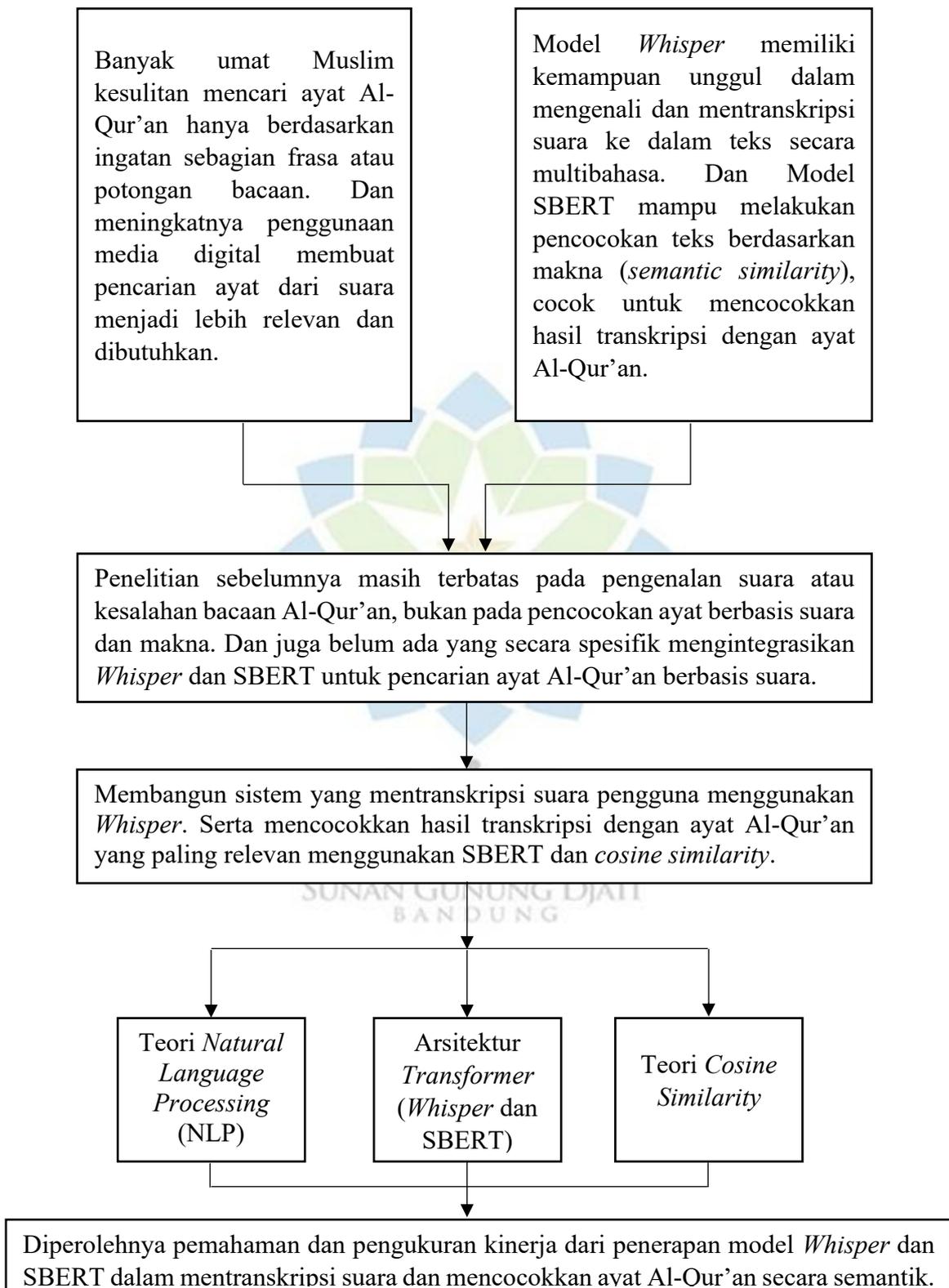
1. Mengimplementasikan model *Whisper* untuk mengubah *input* suara pengguna menjadi teks dengan akurasi yang tinggi.
2. Menerapkan model SBERT untuk mencocokkan hasil transkripsi suara dengan ayat Al-Qur'an berdasarkan kemiripan semantik.
3. Mengukur kinerja model *Whisper* dan SBERT dalam proses pengubahan suara ke teks dan pencocokan ayat secara terpisah.

1.4. Batasan Masalah Penelitian

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

1. Model pengenalan suara yang digunakan adalah *Whisper*, yang digunakan untuk mengubah suara menjadi teks tanpa membandingkannya dengan model lain.
2. Model pencocokan teks yang digunakan adalah *Sentence Bidirectional Encoder Representations from Transformers (SBERT)*, dengan fokus pada pencocokan semantik antara teks hasil transkripsi dan ayat Al-Qur'an tanpa membandingkannya dengan metode pencocokan lainnya.
3. *Dataset* ayat Al-Qur'an yang digunakan berasal dari *platform* Kaggle, mencakup teks dalam bahasa Arab dan terjemahan bahasa Indonesia.
4. *Dataset* suara yang digunakan terbatas pada rekaman suara pengguna yang membaca satu ayat Al-Qur'an secara acak untuk dilakukan proses pencocokan.
5. Penelitian ini tidak membahas aspek tafsir atau pemahaman makna kandungan ayat, melainkan hanya berfokus pada pencocokan berdasarkan kemiripan semantik teks.

1.5. Kerangka Pemikiran Penelitian



Gambar 1.1 Kerangka Pemikiran

1.6. Sistematika Penulisan

Sistematika penulisan skripsi ini disusun agar memudahkan pembaca dalam memahami alur pembahasan penelitian. Adapun sistematika penulisan yang digunakan terdiri dari lima bab utama, yaitu:

- BAB I** : Bab ini berisi latar belakang penelitian, rumusan masalah, tujuan penelitian, batasan masalah, kerangka pemikiran, metodologi penelitian, dan sistematika penulisan. Tujuannya adalah untuk memberikan gambaran umum mengenai topik penelitian, alasan dilaksanakannya penelitian, serta arah dan ruang lingkup penelitian.
- BAB II** : Bab ini memuat tinjauan pustaka dari berbagai penelitian terdahulu yang relevan serta landasan teori yang mendukung penelitian. Teori-teori yang dibahas mencakup Al-Qur'an, Kecerdasan Artifisial (KA), *Deep Learning*, *Machine Learning*, *Natural Language Processing (NLP)*, pengenalan suara, *Whisper*, pencocokan teks, dan *Sentence-BERT (SBERT)*.
- BAB III** : Bab ini menjelaskan metode yang digunakan dalam penelitian, mulai dari pengumpulan data, proses praproses data, arsitektur model *Whisper* dan SBERT, hingga langkah-langkah implementasi serta cara pengukuran kinerja masing-masing model.
- BAB IV** : Bab ini menyajikan hasil implementasi sistem dan pembahasan mengenai kinerja dari masing-masing model yang digunakan. Penjelasan mencakup akurasi transkripsi yang dihasilkan oleh model *Whisper* dan efektivitas pencocokan semantik oleh SBERT dalam mengidentifikasi ayat Al-Qur'an yang relevan berdasarkan *input* suara.
- BAB V** : Bab ini berisi simpulan dari hasil penelitian yang telah dilakukan serta saran untuk pengembangan sistem atau penelitian lebih lanjut yang dapat dilakukan di masa mendatang.