

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi digital, *Internet of Things* atau IoT, dan kecerdasan buatan telah mengubah pola interaksi manusia dengan sistem komputasi. Jika sebelumnya interaksi didominasi antarmuka grafis melalui sentuhan dan teks, kini terjadi pergeseran menuju *Voice User Interface* yang memungkinkan pengguna memberikan perintah atau memperoleh informasi melalui interaksi suara [1].

Interaksi suara dinilai lebih alami karena menyerupai komunikasi antarmanusia, sekaligus mendukung skenario *hands-free* dan *eyes-free*, misalnya saat memasak, mengemudi, atau melakukan aktivitas rumah tangga. Pada asisten suara modern, peningkatan kemampuan pemrosesan bahasa dan integrasi layanan digital turut mendorong meningkatnya pemanfaatan *Voice User Interface* dalam kehidupan sehari-hari, sebagaimana diuraikan oleh [1].

Dalam lingkungan rumah tangga, penggunaan smart speaker memperlihatkan bahwa interaksi suara berlangsung di dalam konteks sosial yang kompleks, misalnya ketika aktivitas lain dan percakapan multipihak terjadi bersamaan seperti pada situasi makan bersama keluarga. Ketika perangkat tidak memberikan respons atau menghasilkan pemahaman yang keliru, pengguna kerap melakukan strategi perbaikan seperti mengulang perintah, melakukan parafrasa, serta menyesuaikan cara pengucapan dan pilihan kata agar sistem merespons, sebagaimana dilaporkan oleh [2].

Dalam penelitian ini, istilah *voice recognition* merujuk pada keberhasilan sistem *Automatic Speech Recognition* (ASR) dalam mengenali perintah suara dan menghasilkan teks yang sesuai dengan maksud perintah, serta tidak dimaknai sebagai pengenalan identitas pembicara. Pada skenario jarak jauh (*far-field*) di lingkungan rumah tangga, pengenalan ujaran menjadi lebih menantang karena sinyal masukan terdegradasi oleh derau latar dan *reverberation* (Boeddeker dkk., 2018.). Selain itu, literatur *far-field* juga menekankan adanya pelemahan sinyal

(*attenuation*) akibat jarak, distorsi ruang, serta gangguan suara yang bervariasi yang dapat menurunkan kualitas sinyal dan menyulitkan proses pengenalan ujaran [4].

Teknologi *voice recognition* pada penelitian ini dipahami sebagai kemampuan sistem ASR mengubah sinyal ujaran menjadi representasi teks yang dapat diproses lebih lanjut untuk kebutuhan perintah, sehingga memungkinkan interaksi manusia–komputer berbasis suara secara otomatis [5]. Salah satu implementasi utama interaksi berbasis suara adalah *smart speaker*, yaitu perangkat konsumen yang memungkinkan pengguna berinteraksi dengan asisten suara untuk mengakses informasi maupun layanan digital melalui perintah suara [1]. Studi empiris juga menunjukkan bahwa asisten suara seperti Alexa, Google Assistant, dan Siri dapat dibandingkan performanya pada tugas tertentu, misalnya dalam memahami ujaran dengan istilah spesifik, sehingga memperkuat relevansi evaluasi sistem berbasis suara pada konteks penggunaan nyata [6].

Meskipun digunakan secara luas, performa sistem pengenalan suara pada *smart speaker* tidak selalu optimal ketika dioperasikan pada kondisi nyata. Faktor lingkungan seperti jarak antara pengguna dan perangkat, derau latar, serta karakteristik akustik ruangan dapat memengaruhi kualitas sinyal suara yang diterima dan berdampak pada akurasi serta respons sistem [5].

Salah satu tantangan utama pada *smart speaker* adalah skenario *far-field speech recognition*, yaitu ketika jarak penutur ke mikrofon relatif jauh. Pada jarak yang meningkat, energi sinyal ujaran menurun (*attenuation*) dan *signal-to-noise ratio* cenderung memburuk, terutama saat terdapat kebisingan latar. Selain itu, pantulan suara di ruangan dapat meningkatkan *reverberation* sehingga pola temporal dan spektral ujaran menjadi kurang jelas. Literatur *far-field* menekankan bahwa kombinasi jarak, kebisingan, dan *reverberation* membuat pengenalan ujaran jarak jauh lebih sulit dibandingkan kondisi mikrofon dekat dan membutuhkan strategi peningkatan ketahanan sistem [5].

Selain faktor akustik, *smart speaker* komersial umumnya mengandalkan pemrosesan berbasis *cloud* untuk menjalankan ASR dan pemahaman bahasa. Arsitektur ini membuat sistem sensitif terhadap kualitas koneksi jaringan; kondisi

seperti latensi tinggi, *jitter*, dan *packet loss* dapat meningkatkan waktu respons atau mengganggu pemrosesan ketika audio tidak terkirim dengan baik ke server [7].

Urgensi penelitian ini terletak pada kebutuhan memastikan keandalan *smart speaker* dalam kondisi penggunaan nyata. Literatur menunjukkan bahwa skenario *far-field* dengan kebisingan dan *reverberation* dapat menurunkan akurasi dan responsivitas pengenalan suara, sementara ketergantungan pemrosesan berbasis *cloud* membuat kualitas jaringan turut memengaruhi waktu respons dan potensi kegagalan pemrosesan [5], [7]. Oleh karena itu, diperlukan evaluasi performa yang terukur untuk memetakan batas kinerja serta faktor dominan yang memengaruhi keberhasilan perintah suara pada kondisi operasional nyata (Boeddeker dkk., 2018).

Amazon Alexa merupakan salah satu platform asisten suara yang banyak digunakan dan kerap dijadikan objek kajian dalam penelitian sistem berbasis suara. Kajian empiris menunjukkan bahwa Alexa dapat dievaluasi performanya pada tugas tertentu dan dibandingkan dengan asisten suara lain, sehingga relevan dijadikan objek evaluasi dalam studi *voice recognition* [6] .

Untuk aspek tantangan pengenalan ujaran jarak jauh pada perangkat rumah tangga, literatur *far-field ASR* memberikan landasan teknis mengenai faktor lingkungan yang memengaruhi performa sistem [5]. Selain *smart speaker* komersial, berkembang pula sistem interaksi suara berbasis *conversational artificial intelligence* yang dirancang sebagai antarmuka komunikasi antara pengguna dan sistem [8].

Dalam penelitian ini, sistem tersebut disebut sebagai *Interactive Speaker System* dan digunakan untuk memungkinkan pengujian performa *voice recognition* secara lebih fleksibel dan terkontrol pada kondisi pengujian yang setara [8]. Perbandingan performa antara *smart speaker* komersial dan *Interactive Speaker System* menjadi penting untuk memperoleh pemahaman yang lebih komprehensif mengenai karakteristik performa *voice recognition* pada arsitektur sistem yang berbeda. Analisis komparatif memungkinkan identifikasi keunggulan dan keterbatasan masing-masing sistem pada kondisi pengujian yang setara, khususnya pada skenario *far-field* yang dipengaruhi oleh jarak dan kondisi lingkungan [5].

Sebagian literatur *far-field ASR* digunakan sebagai rujukan kerangka *robustness* dan praktik evaluasi pada sistem audio, namun penelitian ini berfokus pada evaluasi keberhasilan pengenalan perintah pada *smart speaker* dalam kondisi pengujian yang terdokumentasi [5].

Dengan demikian, performa pada penelitian ini diformalkan sebagai keluaran biner berhasil/gagal berdasarkan kesesuaian makna perintah yang diucapkan dengan hasil pengenalan sistem pada kondisi pengujian yang terdokumentasi, sejalan dengan praktik evaluasi performa pada sistem pengenalan ujaran yang menekankan keterukuran dan reproduisibilitas pengujian [5].

Dalam menganalisis performa *voice recognition*, diperlukan metode analisis yang mampu menangani data multivariat serta hubungan nonlinier antarvariabel. Metode *Random Forest* dipilih karena merupakan pendekatan *ensemble* yang efektif untuk klasifikasi dan relatif tahan terhadap *overfitting*, sehingga sesuai untuk mendukung interpretasi hasil pengujian performa sistem pengenalan suara [9], [10].

Meskipun berbagai kajian telah membahas tantangan *far-field automatic speech recognition* (ASR), seperti pengaruh kebisingan, *reverberation*, dan strategi peningkatan ketahanan sistem, sebagian besar penelitian tersebut berfokus pada pengembangan metode atau arsitektur model dan umumnya dievaluasi pada dataset atau skenario terkontrol [5], [11], [12]. Studi yang secara khusus mengevaluasi performa *voice recognition* pada perangkat *smart speaker* komersial dalam kondisi penggunaan nyata, dengan memadukan variasi jarak pengguna dan dokumentasi kondisi lingkungan secara terukur, masih relatif terbatas jika dibandingkan dengan dominasi studi berbasis benchmark (Porcheron dkk., 2018; Niu dkk., 2025). Selain itu, pendekatan evaluasi yang mengintegrasikan data multimodal serta analisis berbasis pembelajaran mesin untuk menafsirkan faktor dominan keberhasilan pengenalan perintah belum banyak dilaporkan secara komprehensif dalam konteks evaluasi perangkat *smart speaker* apa adanya (Breiman, 2001; Genuer dkk., 2010). Oleh karena itu, diperlukan penelitian yang mengevaluasi performa *voice recognition* secara eksperimental pada perangkat *smart speaker* dalam kondisi

operasional nyata guna memperoleh pemahaman yang lebih utuh mengenai batas kinerja dan faktor-faktor yang memengaruhinya [5], [13].

Berdasarkan uraian tersebut, penelitian ini berjudul “Analisis Performa *Voice Recognition* pada *Smart Speaker* Menggunakan Metode *Random Forest*”, dengan fokus pada analisis komparatif performa *voice recognition* antara Amazon Alexa dan *Interactive Speaker System* pada variasi jarak pengguna dalam kondisi lingkungan yang terdokumentasi [5], [9].

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana performa *voice recognition* pada *smart speaker* Amazon Alexa dan *Interactive Speaker System* ketika digunakan pada kondisi pengujian yang sama?
2. Bagaimana pengaruh variasi jarak pengguna terhadap performa *voice recognition* pada kedua sistem tersebut?
3. Bagaimana metode *Random Forest* digunakan sebagai alat analisis untuk mengklasifikasikan performa *voice recognition* pada *smart speaker*?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, penelitian ini bertujuan untuk mencapai sasaran penelitian sebagai berikut:

1. Menganalisis performa *voice recognition* pada *smart speaker* Amazon Alexa dan *Interactive Speaker System*.
2. Mengevaluasi perubahan performa *voice recognition* berdasarkan variasi jarak pengguna.
3. Menerapkan metode *Random Forest* sebagai alat analisis klasifikasi untuk memperkuat interpretasi hasil pengujian performa *voice recognition*.

1.4 Manfaat Penelitian

Manfaat penelitian ini diharapkan memberikan kontribusi yang terukur dan relevan, sebagaimana diuraikan berikut:

1. Manfaat teoretis berupa penambahan kajian empiris mengenai performa *voice recognition* pada skenario *far-field* berbasis data *multimodal* serta penyajian contoh penerapan metode *Random Forest* untuk analisis klasifikasi dan penafsiran faktor faktor dominan yang memengaruhi keberhasilan pengenalan perintah suara.
2. Manfaat praktis berupa penyediaan gambaran yang terukur mengenai keunggulan dan keterbatasan *smart speaker* komersial dibandingkan sistem pembanding pada variasi jarak pengguna, sehingga dapat menjadi dasar pertimbangan dalam pemilihan perangkat maupun pengembangan solusi interaksi berbasis suara yang lebih andal pada lingkungan nyata.

1.5 Ruang lingkup dan Batasan Penelitian

Ruang lingkup dan batasan penelitian ini ditetapkan sebagai berikut:

1. Penelitian difokuskan pada analisis performa *voice recognition*.
2. Objek penelitian meliputi *smart speaker* Amazon Alexa dan *Interactive Speaker System*.
3. Penelitian tidak mencakup perancangan atau pengembangan perangkat keras *smart speaker*.
4. Analisis performa dilakukan menggunakan metode *Random Forest* sebagai alat klasifikasi.
5. Pengujian smart speaker dilakukan di ruangan non kedap suara untuk merepresentasikan kondisi operasional nyata.