BAB 1. PENDAHULUAN

1.1 Latar Belakang

Pada era teknologi ini yang semakin maju, mesin atau aplikasi pengenalan suara telah menjadi bagian dari kehidupan sehari-hari, baik dalam aplikasi *mobile* maupun perangkat lainnya. Teknologi pengenalan suara memungkinkan pengguna untuk berinteraksi dengan perangkat mereka secara verbal tidak menggunakan interaksi fisik, yang memungkinkan memfasilitasi aksesibilitas yang lebih efisien dan pengalaman pengguna yang lebih baik mulai dari pencarian informasi hingga navigasi. Dari assisten *virtual* seperti *Google Assistant, Amazon Alexa, dan Microsoft Cortana* hingga sistem transkrip otomatis seperti *Amazon Transcribe,* pengenalan suara telah membuka berbagai kemungkinan.

Pengenalan suara telah menjadi semakin penting dalam berbagai aplikasi, seperti pencarian *online*, pengaturan jadwal, dan *control* perangkat dalam rumah. Oleh karena itu, penting untuk memahami kinerja realtif dari model pengenalan suara yang seperti *Google*, terutama dalam konteks penggunaan bahasa Indonesia.

Pengenalan suara atau juga dikenal dengan pengenal ucapan, adalah teknologi yang memungkinkan komputer atau perangkat elektronik unttuk mengenali dan memahami kata-kata yang di ucapkan oleh manusia. Teknologi pengenalan suara berkerja dengan menerjemahkan gelombang suara yang direkam dari ucapan manusia menjadi *text* yang dapat di proses oleh komputer. Teknologi ini semakin canggih seiring waktu. Memungkinkan pengenalan suara yang lebih akurat dan *responsive*. Dengan pengenalan suara pengguna dapat melakukan berbagai tugas tanpa harus mengetik, seperti mencari informasi atau mengirim pesan teks. Ini membawa kenyamanan dan efesiensi yang besar dalam interaksi manusia dan teknologi.

Automatic Speech Recognition (ASR) adalah suatu sistem yang memungkinkan komputer untuk menerima masukan berupa ucapan. Teknologi ini memungkinkan suatu perangkat untuk mengenali dan memahami suara yang diucapkan dengan cara digitalisasi huruf dan mencocokkan sinyal digital tersebut

dengan suatu pola tertentu yang tersimpan dalam suatu perangkat. Hasil dari identifikasi huruf yang diucapkan dapat ditampilkan dalam bentuk tulisan atau dapat dibaca oleh perangkat teknologi sebagai sebuah perintah untuk melakukan suatu pekerjaan(Dyarbirru & Hidayat, 2020).

Penelitian mengenai ASR telah dilakukan selama lebih dari 40 tahun, namun hingga saat ini masih ada penelitian untuk menemukan *Automatic Speech Recognition* yang bisa mengenali ucapan dalam subjek apa pun dengan berbagai bahasa. Salah satu cara untuk meningkatkan tingkat pengenalan adalah dengan menggunakan model dari suatu bahasa itu perlu dikenali (Mustikarini dkk., 2019).

Saat ini, *Mel-frequency cepstral coeffcients* paling umum digunakan dalam pengenalan suara dan verifikasi pembicara karena MFCC dapat bekerja baik pada masukan dengan tingkat korelasi yang tinggi, yaitu dengan menghilangkan informasi yang tidak diperlukan. Selain itu, MFCC bisa mewakili suara manusia dan sinyal music dengan baik karena MFCC menggunakan frekuensi *mel* (Mustikarini dkk., 2019).

Learning vector Quantization (LVQ) adalah teknik dalam pembelajaran mesin yang digunakan unutk mengklasifikasikan data ke dalam kategori atau kelas yang sesauai. LVQ menggunakan serangkaian *vector* prototipe yang mewakili setiap kelas yang mungkin. Selama proses pelatihan, vector-vektor ini disesuaikan agar sesuai dengan pola data pelatihan yang diberikan.

Cara terbaik untuk menguji kualitas beragam dari sistem *Automatic Speech Recognition* (ASR) adalah dengan menggunakan *classification report* sebagai metrik evaluasi utama. Dalam penelitian ini, *classification report* digunakan untuk mengevaluasi kinerja metode MFCC dan LVQ dalam mengenali ucapan. *Classification report* memberikan berbagai metrik penting, seperti *precision, recall, F1-score, dan support,* yang dapat mengGambarkan seberapa baik sistem dalam mengklasifikasikan kata-kata yang diucapkan. *Precision* mengukur seberapa banyak prediksi yang benar dibandingkan dengan seluruh prediksi dalam suatu kelas, sedangkan *recall* menunjukkan seberapa banyak kata dalam kelas tertentu yang berhasil dikenali dengan benar oleh sistem. Sementara itu, *F1-score* adalah

rata-rata harmonik dari *precision* dan *recall*, yang memberikan Gambaran keseimbangan antara keduanya.

Dalam pengenalan suara terdapat beberapa masalah yang harus di ketahui terlebih dahulu. Pertama, variasi suara antar individu, seperti intonasi, tempo, dan kualitas suara, menyebabkan sistem pengenalan suara kesulitan dalam mengenali dan membedakan makna ujaran, terutama dalam membedakan antara ucapan bernada positif dan negatif. Kedua, keragaman dialek dan cara pengucapan dalam bahasa Indonesia turut memengaruhi akurasi sistem dalam mengklasifikasikan makna dari ucapan tersebut. Ketiga, sebagian besar sistem pengenalan suara yang ada belum dioptimalkan untuk mendeteksi kecenderungan emosional atau makna ujaran secara kontekstual, seperti kecenderungan ke arah positif atau negatif. Oleh karena itu, diperlukan metode yang mampu mengekstraksi ciri penting dari sinyal suara dan melakukan klasifikasi secara efektif agar sistem dapat mendeteksi ujaran bermakna negatif (hate speech) maupun positif secara lebih akurat.

Penggunaan *classification report* akan memberikan Gambaran yang lebih komprehensif tentang kinerja model dalam mengenali ucapan bahasa Indonesia. Dengan metrik seperti *precision, recall, dan F1-score*, evaluasi dapat dilakukan secara lebih detail untuk setiap kelas kata yang diuji. Hal ini memungkinkan analisis tidak hanya terhadap kesalahan dalam pengenalan kata, tetapi juga terhadap keseimbangan antara prediksi benar dan salah dalam setiap kategori. Dengan demikian, penelitian ini akan bermanfaat dalam mengembangkan sistem pengenalan ucapan yang lebih akurat dan dapat diandalkan dalam konteks bahasa Indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang sebelumnya, didapatkan rumusan masalah sebagai berikut:

a. Bagaimana proses ekstraksi fitur suara menggunakan metode *Mel-Frequency Cepstral Coefficient* (MFCC) untuk mengidentifikasi karakteristik ucapan positif dan negatif dalam bahasa Indonesia?

b. Bagaimana penerapan metode *Learning Vector Quantizatio*n (LVQ) dalam mengklasifikasikan fitur suara ke dalam kategori ucapan positif dan ucapan negatif?

c. Sejauh mana kombinasi metode MFCC dan LVQ dapat meningkatkan performa klasifikasi ucapan berdasarkan kategori positif dan negatif, ditinjau dari hasil evaluasi menggunakan classification report (*precision, recall, dan F1-score*)?

1.3 Tujuan

Tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk menerapkan metode *Mel Frequency Cepstral Coefficient* (MFCC) dalam mengekstrasi fitur suara untuk mengidentifikasi karakteristik ucapan positif dan negatif dalam bahasa Indoneisa.
2. Untuk menerapkan metode *Learning Vector Quantization* (LVQ) dan parameter yang sudah ditentukan dalam mengklasifikasi fitur suara hasil ekstrasi ke dalam kategori ucapan positif dan negatif.
3. Untuk mengevaluasi peforma model klasifikasi dari kombinasi metode MFCC dan LVQ dengan menggunakan metrik *Classification Report* yang mencakup *precision, recall,* dan *F1-Score.*

1.4 Manfaat

Penelitian ini memiliki potensi yang mendukung dalam pengembangan aplikasi yang berbasis pengenalan suara yang lebih efektif dan efisien dalam berbagai konteks, meningkatkan kualitas layanan dan meningkatkan pengalaman pengguna secara keseluruhan

1.5 Batasan Masalah

Adapun batasan masalah yang dilakukan dalam penelitian ini yaitu:

a. Penelitian ini akan difokuskan pada penggunaan metode *Mel-Frequency Cepstral Coeffcient* (MFCC) dan teknik *Learning Vector Quantization* (LVQ) dalam implementasi model pengenalan suara ucapan positif dan

negatif untuk bahasa Indonesia, dengan tidak melibatkan teknik atau metode lain dalam analisis.

b. Penelitian ini akan difokuskan pada pengembangan dan evaluasi sistem pengenalan suara berbasis MFCC dan LVQ tanpa membandingkannya dengan platform pengenalan suara komersial lainnya. Evaluasi kinerja akan dilakukan tanpa mempertimbangkan aspek lain seperti kecepatan pengenalan atau fitur tambahan.

c. Penelitian ini akan membatasi pengujian dan evaluasi terhadap data ucapan dalam bahasa Indonesia yang telah terverifikasi dan terstruktur, tanpa memperluas cakupan pada bahasa lain atau data ucapan yang belum diverifikasi.