

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Dengan semakin berkembangnya teknologi telekomunikasi, internet menjadi sesuatu yang tidak lagi sulit dan mahal. Kemudahan ini menyebabkan internet dipenuhi berbagai macam informasi yang tidak hanya berupa teks, tetapi juga gambar, audio, dan video. Kebutuhan akan identifikasi pembicara juga semakin bertambah, sehingga perolehan informasi mengenai seseorang akan dapat menghasilkan data audio atau video dari orang tersebut. Misalnya, seorang wartawan ingin mencari seseorang dari data suara yang ada, dengan identifikasi pembicara hal itu dapat dilakukan. Data audio dan video yang diperoleh dapat memperkaya informasi mengenai orang tersebut sehingga tulisan yang akan dibuat oleh si wartawan menjadi semakin menarik.

Identifikasi pembicara (*speaker identification*) merupakan suatu proses untuk mengidentifikasi siapa yang berbicara (pembicara) pada suatu sesi data *speech*. Pada umumnya, identifikasi pembicara terbagi atas *text independent* dan *text dependent*. Untuk kategori *text independent* proses identifikasi tidak dipengaruhi oleh isi pembicaraan si pembicara. Sebaliknya, pada *text dependent*, seluruh pembicara harus menggunakan kata yang sama atau beberapa kata yang sudah ditentukan. Selain itu, berdasarkan data yang diproses, identifikasi pembicara terbagi atas dua, yaitu berbasis linguistik dan berbasis sinyal akustik. Jika berbasis linguistik, data akan berisi unsur kebahasaan, seperti kata, kalimat, dan kamus data. Sedangkan yang berbasis sinyal akustik, data akan merepresentasikan bagaimana karakteristik *speech* tersebut terdengar di telinga manusia. Sistem yang berbasis sinyal akustik tidak mempunyai batasan kebahasaan.

Aplikasi identifikasi pembicara banyak digunakan untuk keamanan, yaitu biometrik, dan perolehan informasi multimedia. Pada aplikasi biometrik, data pembicara yang digunakan biasanya tidak sebanyak pada aplikasi sistem perolehan informasi dan lebih mengarah ke verifikasi pembicara dari pada

identifikasi pembicara. Sedangkan sistem perolehan informasi berkembang karena adanya kebutuhan pencarian data tertentu pada koleksi data yang semakin banyak.

Penelitian di bidang identifikasi pembicara sudah berlangsung lama. Hampir semua metode sudah digunakan. Metode yang dijadikan acuan adalah Gaussian Mixture Models (GMM) [Reynolds dan Rose, 1995] yang menghasilkan akurasi yang sangat baik, yaitu 80,8% untuk data *speech* dengan kualitas pembicaraan lewat telpon (*telephone speech*) dan 96,8% untuk data *speech* yang jernih (*clean speech*). Metode GMM digunakan berdasarkan interpretasi bahwa komponen Gaussian dapat merepresentasikan bentuk spektral pembicara secara umum dan kemampuan Gaussian mixtures dalam memodelkan densitas data. Metode lain yang digunakan adalah Hidden Markov Model (HMM) [Buono et al, 2008]. Metode ini memodelkan tidak hanya suara, tapi juga urutan suara, dan lebih cocok untuk *text-dependent speaker identification*. Selain itu, metode Vector Quantization (VQ) [Kamruzzaman et al, 2007] juga sudah digunakan. Dengan metode ini, setiap pembicara mempunyai *codebook* berisi model spektral dari suaranya. Yang terakhir adalah Support Vector Machine (SVM) [Schmidt dan Gish, 1996]. Metode SVM digunakan karena sederhana dan lebih efektif, di mana fitur suara dari pembicara dapat digunakan secara langsung tanpa perlu estimasi densitas terlebih dahulu.

Berdasarkan data yang diproses, penelitian identifikasi pembicara terbagi atas yang berbasis linguistik dan sinyal akustik. Untuk sinyal akustik sendiri, fitur data yang digunakan antara lain berupa Linear Predictive Coding (LPC) [Wan dan Campbell, 2000], Mel-scale Frequency Cepstrum Coefficient (MFCC) [Moreno dan Ho, 2008], Mel-scale Bispectrum [Buono et al, 2008], dan wavelet [Lin et al, 2006]. MFCC banyak digunakan karena dianggap yang paling sesuai dalam memodelkan frekuensi suara manusia. Sedangkan bispektrum digunakan untuk mengatasi *noise* pada suara, namun hal ini bisa juga diatasi dengan menggunakan principal component analysis (PCA) pada saat ekstraksi fitur suara [Li et al, 2008].

Tesis ini mengusulkan sistem identifikasi pembicara dengan menggunakan metode klasifikasi SVM dan fitur audio MFCC. Data *speech* yang digunakan

adalah data berbahasa Indonesia yang berasal dari beberapa stasiun berita televisi dan radio.

1.2 Perumusan Masalah

Berdasarkan latar belakang di atas, dapat dirumuskan beberapa permasalahan sebagai berikut:

1. Mengetahui hasil identifikasi pembicara dengan metode SVM.
2. Mengetahui panjang data dan jumlah data yang cukup dari setiap pembicara untuk dapat memperoleh hasil identifikasi yang cukup akurat.
3. Mengetahui metode SVM *multi-class* yang tepat untuk klasifikasi kelas yang sangat banyak sehingga dapat diperoleh hasil identifikasi yang baik.
4. Mengetahui metode pemrosesan data suara untuk dapat menghilangkan gangguan pada suara (*noise*).

1.3 Tujuan Penelitian

Tesis ini bertujuan untuk melakukan identifikasi pembicara pada data berita berbahasa Indonesia dengan menggunakan salah satu metode *machine learning*, yaitu SVM. Fitur suara yang digunakan adalah MFCC.

1.4 Ruang Lingkup Penelitian

Ruang lingkup penelitian dalam tesis ini adalah sebagai berikut:

1. Koleksi data yang digunakan dalam tesis ini berasal dari stasiun berita televisi dan radio, di mana suara yang diperoleh masih mengandung *noise*.
2. Karena keterbatasan sumber daya komputasi, klasifikasi tidak akan menggunakan seluruh frame suara yang ada melainkan hanya rata-rata nilai frame dalam suatu waktu, yaitu 5, 10, dan 15 detik.
3. Jumlah data untuk setiap pembicara tidak sama (*unbalanced data*).

1.5 Metodologi Penelitian

Penelitian ini dilakukan dengan langkah-langkah sebagai berikut:

1. Perumusan Masalah dan Studi Literatur

Pada tahap ini akan dilakukan analisa terhadap masalah, pengumpulan bahan-bahan dan referensi, untuk dijadikan bahan acuan dalam melakukan studi awal pemahaman konsep dan perumusan model sistem yang akan dibuat. Literatur-literatur ini diperoleh melalui penelusuran jurnal, makalah, buku, dan informasi lain yang terkait dengan tesis ini.

2. Pengumpulan Data

Data yang digunakan berasal dari data berita berbahasa Indonesia dari beberapa stasiun radio dan televisi yang ada.

3. Perancangan Sistem

Pada tahap ini, dilakukan perancangan sistem identifikasi pembicara yang menggunakan metode SVM dan fitur suara MFCC.

4. Implementasi Sistem

Tahap ini merupakan implementasi dari tahap perancangan sistem. Sistem yang dibuat terbagi atas pra-pemrosesan, ekstraksi fitur *speech*, pemodelan pembicara, dan identifikasi pembicara.

5. Uji Coba Sistem

Pada tahap ini dilakukan identifikasi dari sejumlah data *speech*. Hasil evaluasi akan dianalisa kebenarannya.

1.6 Sistematika Penulisan

Tesis ini disusun dengan sistematika sebagai berikut:

Bab 1 PENDAHULUAN: berisi penjelasan mengenai latar belakang, rumusan masalah, tujuan penelitian, ruang lingkup penelitian, metodologi, serta sistematika penyusunan laporan penelitian.

Bab 2 LANDASAN TEORI: membahas algoritma, metode, dan teori-teori yang digunakan dalam perancangan sistem identifikasi pembicara.

Bab 3 EKSPERIMEN: berisi penjelasan mengenai eksperimen yang dilakukan, yaitu: data yang digunakan, perancangan sistem, implementasi sistem, dan pengujian sistem.

Bab 4 HASIL UJICOBA DAN ANALISA: berisi hasil uji coba sistem yang telah dikembangkan dan analisa unjuk kerja sistem identifikasi pembicara berdasarkan hasil pengujian yang diperoleh.

Bab 5 KESIMPULAN DAN SARAN: berisi kesimpulan dari hasil penelitian ini dan saran untuk pengembangan selanjutnya.

