

BAB I

Pendahuluan

1.1. Latar Belakang

Internet merupakan sarana pertukaran data yang kuat. Data dapat tersirkulasi secara cepat dan dapat diakses dengan mudah. Penggunaan internet tumbuh secara eksponensial dari jumlah pengguna maupun jumlah datanya. Salah satu jenis data tersebut berupa video.

Penggunaan video sebagai media semakin marak. Banyaknya platform yang menyediakan layanan berbagi video menyebabkan siapapun dapat mengunggah video di Internet. Bahkan jaringan televisi besar pun menggunakan Internet untuk sarannya. Salah satu platform berbagi video terbesar adalah Youtube. Youtube memiliki miliaran pengguna, hampir sepertiga dari pengguna internet dan setiap harinya pengguna melihat ratusan juta jam video [1]. Bahkan sekitar 500 jam video yang diunggah pada Youtube setiap menitnya pada tahun 2015 [2]. Pada tahun 2015, 70% dari pengguna internet mengakses video dan pada tahun 2020 diprediksikan akan mencapai 82% [3].

Dengan jumlah data dan pengguna yang besar tersebut, maka jumlah informasi yang dikandungnya signifikan. Apabila informasi tersebut dapat diolah, maka hasil olahan tersebut dapat digunakan untuk berbagai kebutuhan, salah satu contohnya adalah iklan. Dapat dilakukan analisis pola pada video yang populer untuk membuat iklan dengan tema yang populer tersebut.

Hidden Markov Toolkit (HTK) dikembangkan oleh Cambridge University Engineering Department (CUED). HTK ditujukan untuk membangun dan memanipulasi Hidden Markov Model (HMM). Kegunaan utama HTK untuk penelitian pengenalan ucapan meskipun dapat digunakan pada apapun yang menggunakan *data time series* [4].

Pada tahun 2010, Janakiraman dkk. menunjukkan pengenalan ucapan berdasarkan silabel mampu menurunkan kompleksitas perhitungan dan *word error rate*(WER) menjadi 4,4% pada database TIMIT dan 21,2% pada database NTIMIT [5]. Tahun 1994, Mak dan Allen *audio-visual speech segmentation system* berdasarkan kombinasi sinyal akustik dari ucapan dan sinyal visual dari bibir berhasil menurunkan *segmentation error* lebih dari 10,4% terutama pada sinyal ucapan yang rusak karena *noise* [6]. Diikuti dengan Talea dan Yaghmaie ekstraksi bibir dan segmentasi silabel pada bahasa Persia yang menunjukkan hasil yang menjanjikan [7].

1.2. Perumusan Masalah

Pada data video, akan banyak ditemukan pengucapan kontinu. Sehingga untuk pengolahan ucapan pada data video, dengan menggunakan sinyal ucapannya saja, maka masih tersisa data yang dapat dimanfaatkan.

1.3. Batasan Masalah

Adapun yang menjadi batasan masalah pada penelitian ini adalah:

1. Sistem multimodal pengenalan ucapan dibatasi hanya dengan pergerakan bibir saja.
2. Sistem dibangun menggunakan HTK *toolkit*.
3. Sistem yang dibangun merupakan *speaker dependant system*
4. Data yang direkam dari satu orang mengucapkan 5 kalimat dengan noise yang sudah dibersihkan.

1.4. Tujuan

Membangun sistem pengenalan ucapan kontinu berbasis silabel dengan bantuan pengenalan visual pada pergerakan bibir .

1.5. Metodologi

Berikut adalah metodologi yang dilakukan untuk menyelesaikan masalah yang muncul.

1.5.1. Studi Literatur

Mengumpulkan dan mempelajari kajian yang mampu mendukung dalam pemecahan masalah yang muncul pada penelitian ini.

1.5.2. Analisis Kebutuhan Sistem

Dilakukan analisis terhadap kebutuhan sistem untuk mencapai tujuan pada tugas akhir ini.

1.5.3. Pengujian Sistem

Melakukan proses *training* dan evaluasi pada sistem lalu dibuat analisis sementara.

1.5.4. Analisis

Menganalisis hasil output dari sistem. Menghitung *word error rate* dan akurasi berdasarkan hasil evaluasi.

1.5.5. Penyusunan Laporan

Pembuatan laporan mengenai kegiatan dan sistem yang dibangun yang meliputi latar belakang, rumusan masalah, tujuan, implementasi sistem hingga hasil analisis yang dilakukan selama pengerjaan tugas akhir