

# 1. Pendahuluan

Bagian ini berisi penjelasan latar belakang dipilihnya permasalahan penelitian, perumusan masalah, tujuan, metodologi, dan sistematika penulisan tugas akhir.

## 1.1 Latar Belakang

Bahasa Arab merupakan bahasa yang unik karena bahasa Arab sangat memperhatikan *makhraj* (tempat keluarnya huruf) yang menjadi pembeda antar huruf dan kata. Bahasa Arab yang merupakan bahasa pembangun kitab Al-Qur'an, merupakan bahasa yang penting untuk dipelajari. Al-Qur'an merupakan buku suci yang dibaca oleh kurang lebih 1.5 juta pengikut umat Islam di dunia [1]. Ada beberapa jenis pengucapan dalam lafal Al-Qur'an, diantaranya adalah *tartil* (pembacaan lambat), *tadwir* (sedang) dan *hadr* (cepat) [2]. Perbedaan pengucapan ini kadang berujung pada perbedaan makna kata karena pengucapan Al-Qur'an sangatlah terkait dengan *harakat* (panjang pendeknya suatu kata). Selain itu pengucapan kata yang berbeda (logat bahasa) juga dapat mempengaruhi arti dari kata tersebut [3]. Dikarenakan masalah tersebut, maka sistem rekognisi *speech* terhadap huruf Hijaiyah berharakat dalam Al-Qur'an sangatlah diperlukan sebagai salah satu sarana pembelajaran bahasa Arab.

Ada banyak metode yang dapat digunakan untuk membangun sistem rekognisi *speech* huruf Hijaiyah berharakat dalam Al-Qur'an. Salah satu metode yang paling menonjol untuk melakukan rekognisi terhadap *speech* adalah *Hidden Markov Model* atau biasa disebut HMM [4]. HMM merupakan metode yang sangat sering digunakan di dalam berbagai bidang, seperti: *image analysis*, *speech recognition*, *handwritten recognition* dan *signal processing* [5]. *Inference* utama yang digunakan dalam HMM adalah *Bayes' Rule*, yang mana digunakan juga dalam *Naïve Bayes* [4] yang merupakan salah satu bagian dari *Bayesian Network*. Oleh karena itu, sistem ini akan lebih memfokuskan penggunaan dari *Naïve Bayes* dan *Bayesian Network* dalam hal rekognisi terhadap data *speech*. Sebelum proses rekognisi dilakukan, data *speech* akan melewati tahap *pre-processing* menggunakan *Linear Predictive Coding* atau biasa disebut LPC, guna mencari koefisien *cepstral* yang kemudian akan digunakan sebagai data inputan *classifier*.

Sistem rekognisi ini diharapkan dapat membantu orang khususnya yang beragama Islam dalam pembelajaran bacaan Al-Qur'an yang benar. Selain itu, data

set dari sistem rekognisi ini pun dapat dijadikan data set baru demi perkembangan sistem rekognisi bahasa Arab khususnya Al-Qur'an.

## 1.2 Perumusan Masalah

Rumusan Masalah yang diangkat pada tugas akhir ini adalah

- a. Bagaimana cara mengklasifikasi huruf Hijaiyah yang berada di dalam Al-Qur'an?
- b. Bagaimana hasil analisis dari pendeteksian ucapan huruf Hijaiyah yang berada di dalam Al-Qur'an?

## 1.3 Tujuan

Berdasarkan perumusan masalah diatas maka tujuan yang ingin dicapai adalah

- a. Membangun sistem yang dapat mengklasifikasikan ucapan huruf Hijaiyah bertanda baca dengan menggunakan metode *Naïve Bayes* dan *Bayesian Network*
- b. Menganalisis performansi sistem pendeteksi ucapan huruf Hijaiyah bertanda baca yang diajukan berdasarkan *micro-average F-1 score*

## 1.4 Metodologi

Metodologi pada pembuatan laporan tugas akhir ini adalah

### a. Studi Literatur

Studi literatur dilakukan dengan cara *mereview* paper-paper yang berkaitan dengan topik yang diangkat pada tugas akhir ini, yaitu tentang konsep *Naïve Bayes* dan *Bayesian Network*.

### b. Pengumpulan Data

Dataset direkam dari 6 orang yang terbagi atas 3 laki-laki dan 3 perempuan. Data suara dengan pengucap laki-laki memiliki karakteristik pengucapan *hadr* (cepat), sedangkan data suara dengan pengucap perempuan memiliki karakteristik pengucapan *tartil* (lambat). Setiap orang membaca semua huruf Hijaiyah dengan semua tanda baca yaitu *fathah*, *kasrah*, *dhammah*, *fathahtain*, *kasrahtain*, dan *dhammahtain*. Data direkam dengan menggunakan alat bantu rekam dan disimpan kedalam file dengan format *wav*.

c. Perancangan Sistem

Merancang sistem dengan metode *Naïve Bayes* dan *Bayesian Network*.

d. Implementasi Sistem

Mengimplementasikan sistem dengan cara menggunakan data *training* sebagai data inputan kedalam sistem.

e. Pengujian Sistem

Melakukan pengujian sistem dengan menggunakan data *testing*. Sistem akan mengklasifikasikan data *testing* dan memberikan label terhadap data tersebut.

f. Analisis Sistem

Melakukan observasi dan pengujian terhadap sistem yang telah diimplementasikan dan kemudian dianalisis.

g. Penyusunan Laporan Tugas Akhir

Membuat laporan tugas akhir yang menjelaskan keseluruhan sistem yang telah diimplementasikan, diuji, dan dianalisis.

## 1.5 Sistematika Penulisan

Laporan tugas akhir ini disusun dengan sistematika penulisan sebagai berikut:

1. Pendahuluan

Bab ini berisi penjelasan latar belakang dipilihnya permasalahan penelitian, perumusan masalah, tujuan, metodologi penyelesaian masalah dan sistematika penulisan.

2. Landasan Teori

Bab ini berisi landasan teori yang mendukung penelitian pada tugas akhir ini. Landasan teori meliputi penjelasan mengenai *speech recognition*, *Arabic speech recognition*, metode yang dipakai untuk proses pengubahan data suara ke dalam bentuk *cepstral coefficient* yaitu *Linear Predictive Coding*, *k-fold cross validation*, *Principal Component Analysis*, *Bayesian Network* serta *Naïve Bayes* yang digunakan untuk melakukan pengklasifikasian data dan *F-1 measure* yang digunakan untuk mengukur performansi sistem.

3. Perancangan Sistem

Bab ini memaparkan analisis kebutuhan sistem dan rancangan sistem yang digunakan, serta tahapan implementasi sistem.

4. Pengujian dan Analisis

Bab ini menguraikan hasil pengujian terhadap sistem yang dibuat beserta hasil analisis dari pengujian yang telah dilakukan.

5. Kesimpulan dan Saran

Bab ini berisi kesimpulan dari hasil pengujian dan analisis yang telah dilakukan serta saran untuk penelitian kedepannya.