

INTISARI

Riset mengenai ASR telah dilakukan selama lebih dari 40 tahun, namun hingga saat ini masih dikaji untuk menemukan ASR yang dapat mengenali tutur dalam subyek apapun oleh bermacam-macam bahasa. Salah satu cara untuk meningkatkan *recognition rate* adalah dengan menggunakan model dari bahasa yang ingin dikenali. Bahasa Indonesia sebagai salah satu bahasa *non-mainstream* belum memiliki korpus tutur yang mumpuni dibanding bahasa lain, sehingga berpengaruh pada tingkat pengenalan tutur dalam Bahasa Indonesia.

Pada penelitian ini, dikembangkan suatu sistem pengenalan tutur untuk mengenali kata “atas”, “bawah”, “kanan” dan “kiri”. Data tutur didapatkan dari akuisisi suara secara *offline* untuk data acuan dan secara *real-time* untuk data uji. Data yang digunakan adalah 400 sampel suara, 300 sampel untuk data latih dan 100 sampel untuk data uji yang terdistribusi rata dalam empat kelas.

Sistem pengenalan tutur menggunakan *mel frequency cepstral coefficient* (MFCC) sebanyak 13 koefisien per sampel sebagai ciri dan *support vector machine* (SVM) sebagai pengenalan. Untuk mengoptimasi sistem pengenalan tutur, pengenalan dilatih dengan tiga dataset yang berisi 25 sampel per kelas, 50 sampel per kelas dan 75 sampel per kelas, dan pengenalan diuji dengan data uji. Penggunaan 75 sampel per kelas menghasilkan rata-rata nilai presisi sistem terbaik sebesar 100%.

Setelah itu dengan menggunakan 75 sampel per kelas, *hyperparameter* berupa kernel dan nilai *cost* dari *classifier* SVM diubah-ubah dengan pencarian *grid* menggunakan *10-folds cross validation*. Dari pencarian *grid*, didapatkan *hyperparameter* terbaik berupa kernel linier dan nilai *cost* sebesar 1000 atau $C = 1000$.

Dengan data latih 75 sampel per kelas serta *hyperparameter* SVM berupa kernellinier, nilai *cost* sebesar 1000, sistem menghasilkan nilai F1, yaitu nilai *trade-off* antara *precision* dan *recall*, pada tahap pelatihan sebesar 80% untuk kata “atas”, 81% untuk kata “bawah”, 86% untuk kata “kanan” dan 100% untuk kata “kiri”. Sedangkan pada tahap pengujian dihasilkan nilai F1 sebesar 76% untuk kata “atas”, 44% untuk kata “bawah”, 54% untuk kata “kanan” dan 100% untuk kata “kiri”.

Kata kunci: *automatic speech recognition*, Bahasa Indonesia, MFCC, SVM, *machine learning*.

ABSTRACT

Research about ASR has been going around for more than four decades, but it is still the main objective to be able to recognize speech of any subjects in any languages. One way to increase the recognition rate is to use the language model of the said language. Bahasa Indonesia as one of the non-mainstream languages doesn't have a robust model yet, hence it impacts the recognition rate of speech in Bahasa Indonesia.

This research proposed a design of a speech recognition system that recognizes the word "atas" (means up), "bawah" (means down), "kanan" (means left), and "kiri" (means right). Speech data was acquired offline for reference data and real-time for testing data. The system used 400 samples, 75 samples from each word for training data and 25 samples from each word for testing data.

The system used mel frequency cepstral coefficient (MFCC) with 13 coefficients per sample as its input features and support vector machine (SVM) as its classifier. The classifier was trained using three datasets consisted of 25 samples per class, 50 samples per class and 75 samples per class. The classifier was then tested using the testing data. The outcome showed that using 75 samples per class yield the best result, with the average precision score of all classes was 100%.

Using 75 samples per class, some SVM hyperparameters, such as kernels, cost and gamma values, were tuned using grid search with 10-folds cross validation to get the best hyperparameters. Grid search showed that the best system performance was achieved using a linear kernel and a cost value of 1000 or $C = 1000$.

Using 75 samples per class, a linear kernel and a cost value of 1000, the system showed an f1-score (the trade-off score between precision and recall) of 80% for "atas", 81% for "bawah", 86% for "kanan" and 100% for "kiri" during the training phase. Meanwhile, in testing phase, the system showed an f1-score of 76% for "atas", 44% for "bawah", 54% for "kanan" and 100% for "kiri".

Keywords: *automatic speech recognition, Bahasa Indonesia, MFCC, SVM, machine learning*