

1. Pendahuluan

Latar Belakang

Identifikasi Bahasa (*Language Identification*) adalah salah satu *task* dalam Pemrosesan Bahasa Alami (*Natural Language Processing*) yang bertujuan mengenali sebuah bahasa yang ada di dalam sebuah teks atau dokumen.[1] Identifikasi Bahasa dapat diaplikasikan di berbagai area, proses, atau *task* seperti *document resume*, *text classification*, *machine translation*, dan lainnya. Salah satu contoh penggunaan dari sistem identifikasi bahasa adalah pengembangan lebih lanjut dari *chatbot* dimana sistem secara otomatis mengidentifikasi bahasa dari lawan bicara sebelum memberi respon, analisis sentimen pada kalimat.

Saat ini penelitian terhadap *Language Identification* mengalami perkembangan yang pesat. Terdapat beberapa hasil penelitian yang membahas identifikasi Bahasa diantaranya *labeling hate speech*,[2] dan mengidentifikasi bahasa isyarat.[3] Namun untuk bahasa dengan korpus yang terbatas atau bahasa yang minor digunakan dalam penelitian, relatif sulit untuk menemukan hasil penelitian yang membahas tentang identifikasi bahasa tersebut, salah satu bahasa yang dimaksud adalah bahasa daerah yang terdapat pada di Indonesia, adapun saat ini penelitian identifikasi bahasa terhadap bahasa daerah yang dilakukan adalah *spoken language identification* dan identifikasi bahasa daerah di Indonesia dengan berbagai metode *machine learning*.

Topik dan Batasannya

Masalah yang diangkat dari Tugas Akhir ini adalah bagaimana mengenali sebuah bahasa dari input kalimat yang berupa teks dengan bahasa daerah tertentu. Adapun bahasa daerah yang ada di Indonesia seperti bahasa Jawa, Sunda, Banjar, Bali, dan lainnya. Untuk Batasan pada Tugas Akhir ini Bahasa yang digunakan dalam eksperimen adalah Bahasa Indonesia, Bahasa Jawa, Bahasa Sunda, Bahasa Minang, Bahasa Muna, Bahasa Madura, dan Bahasa Bugis.

Tujuan

Tujuan dari Identifikasi Bahasa daerah pada Tugas Akhir ini adalah untuk membuat identifikasi Bahasa daerah dengan menggunakan metode multinomial naïve bayes. Aplikasi ini lebih lanjut dapat digunakan untuk *user profiling*, mesin translasi, dan lainnya.

Organisasi Tulisan

Laporan tugas akhir ini dituliskan dengan sistematika sebagai berikut: Pendahuluan untuk menjelaskan latar belakang serta batasan penelitian dan tujuan dari tugas akhir ini. Selanjutnya dijelaskan studi terkait materi yang mendukung tugas akhir yakni studi tentang *language identification* dengan perkembangannya terhadap bahasa daerah, Naïve bayes dan. Setelahnnya Sistem Identifikasi bahasa daerah, analisis hasil, serta kesimpulan.