

SARI

Indonesia merupakan negara yang memiliki jumlah bahasa etnis terbesar kedua di dunia. Menurut Ethnologue, lembaga bahasa di dunia, Indonesia memiliki 707 bahasa daerah. Negara dengan jumlah bahasa etnis paling banyak adalah Papua Nugini, dengan jumlah 839 bahasa etnis. Namun, tidak sedikit dari bahasa daerah di Indonesia yang terancam, bahkan sudah punah. Beberapa diantara penyebab punahnya bahasa daerah adalah faktor urbanisasi, perkawinan antar etnis, penggunaan bahasa daerah dalam pendidikan, serta bahasa dominan dalam suatu wilayah multibahasa yang berdampingan.

Salah satu wadah yang memungkinkan untuk menjadi media pendekatan kepada generasi muda saat ini adalah melalui sosial media. Hal ini disebabkan oleh maraknya penggunaan sosial media oleh seluruh kalangan di Indonesia khususnya generasi muda. Sehingga diperlukan suatu wadah untuk mengakses sosial media yang bisa menampilkan tulisan dari pengguna yang ada di sosial media tersebut dan menampilkan bahasa daerah mana yang digunakan dalam penulisan tersebut.

Penelitian dilakukan dengan menggunakan data dari Twitter, salah satu sosial media yang berbasis teks, yang mana hanya mendeteksi bahasa dari negara tertentu dan belum bisa menentukan bahasa dari daerah-daerah yang spesifik seperti bahasa daerah di Indonesia. Dan hasil dari penelitian ini merupakan model klasifikasi bahasa yang bisa digunakan untuk mendeteksi bahasa daerah, khususnya bahasa Jawa dan Sunda. Serta juga hasil dari pengujian dari model yang telah dibuat, sehingga dapat ditentukan bahwa model sudah cukup layak untuk digunakan.

Kata kunci: Twitter, klasifikasi, sosial media, bahasa daerah, model.