

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Indonesia adalah negara yang terdiri dari berbagai macam suku, etnis, bahasa, agama, dan peradaban. Keanekaragaman ini merupakan keunggulan yang tak tertandingi yang tidak dimiliki oleh bangsa lain. Praktik-praktik kebahasaan dari seorang individu atau kelompok dapat mencerminkan perilaku berbahasa masyarakat Indonesia. Penerapan bahasa Indonesia yang akurat dan tepat bergantung pada konteks, situasi, dan ranah kehidupan yang beragam. Keragaman bahasa secara intrinsik terkait dengan budaya, karena bahasa berfungsi sebagai komponen fundamental budaya, seperti yang ditunjukkan oleh bahasa Lampung. Bahasa Lampung diakui sebagai bahasa daerah.

Penelitian sebelumnya menunjukkan bahwa Lampung adalah sebuah daerah di pulau Sumatra yang memiliki bahasa daerah dan aksara tersendiri. *Stemming* kata-kata bahasa Lampung dalam dialek Api dilakukan dengan menggunakan pendekatan *brute-force*. Bahasa Lampung memiliki kata kerja infleksional (berasal dari kata dasar verba) dan kata kerja derivasional (berasal dari kata dasar dalam kategori selain verba, termasuk kata benda, kata sifat, kata keterangan, kata ganti, dan kata bilangan). Penelitian ini bertujuan untuk mengimplementasikan *stemming* kata melalui metode *brute-force*, mengembangkan aplikasi stemming kata dialek Lampung dalam bahasa C#, dan memanfaatkan basis data online melalui *Firebase*. Teknik ini melibatkan

peneliti untuk mencari, membuat katalog, dan memasukkan 2000 kata tetap dari istilah inti dialek Lampung 'api' secara manual, mengembangkan algoritma stemming dengan metode *brute-force*, dan mengevaluasi aplikasi yang dihasilkan. Temuan penelitian menunjukkan bahwa aplikasi ini dapat melakukan stemming kata untuk 2000 istilah yang telah ditentukan. Dengan tidak adanya stemming yang layak, sebuah metode tersedia untuk menyegarkan basis data yang digunakan oleh program, sehingga meningkatkan kemampuan stemming-nya, yang bermanfaat untuk aplikasi penerjemahan mesin berbasis kamus (Abidin et al., 2021).

Penelitian sebelumnya terhadap istilah “api” dalam dialek Lampung menggunakan pendekatan *brute-force*, di mana prosedur stemming melibatkan 2000 kata berimbuhan dan bentuk dasarnya, yang diimplementasikan dengan menggunakan bahasa pemrograman C#. Metodologi ini menyempurnakan mesin penerjemah berbasis kamus untuk bahasa Lampung, khususnya untuk istilah-istilah yang tidak ada dalam basis data aplikasi. Metode *brute-force* untuk stemming bahasa Lampung memerlukan penyandingan kata masukan dengan entri dalam basis data kata yang sesuai. Teknik *brute-force* ini dibatasi oleh ketiadaan kata berimbuhan di antara 2000 kata berimbuhan yang terhubung dengan kata dasar.

*Brute-force* dapat digunakan untuk mengembangkan mesin penerjemah frasa untuk API dialek Lampung yang secara efektif menangani kata-kata berimbuhan (Abidin et al., 2021), yang memfasilitasi penyalarsan kata-kata yang diinput dengan kata-kata yang ada di database kata berimbuhan. Namun, *brute-force* terbatas untuk memproses hanya 2000 istilah berimbuhan yang ditentukan dan bentuk dasarnya.

Penelitian ini menggunakan Algoritma Nazief & Adriani untuk melakukan stemming pada bahasa Lampung. Bahasa Lampung memiliki keunikan tersendiri, terutama dalam hal *morfologi*. *Morfologi* adalah disiplin ilmu yang mempelajari unsur-unsur linguistik, termasuk satuan-satuan gramatikal. Proses *morfologi* melibatkan pembentukan kata dari bentuk dasar melalui proses *afiksasi*, *reduplikasi*, dan *fusi*. Tingkat variasi morfologis dalam suatu bahasa sangat mempengaruhi bahasa tersebut. Oleh karena itu, kata-kata yang memiliki beberapa varian morfologi harus diganti dengan bentuk dasarnya.

Penelitian ini bertujuan untuk meningkatkan leksikon Bahasa Lampung dengan menggunakan versi modifikasi dari Algoritma Nazief & Adriani. Teknik stemming digunakan untuk menstandarisasi teks dan meningkatkan pemrosesannya. Ini adalah fase penting dalam pra-pemrosesan teks. Program stemming biasanya disebut sebagai algoritma stemming atau stemmer. Hasil yang diharapkan dari metodologi Nazief & Adriani untuk pengguna adalah peningkatan pengenalan kata dasar dalam istilah berimbunan Bahasa Lampung, yang dicapai melalui dua penelitian untuk mengevaluasi nilai *Gold Standard Assessment*

## **1.2 Rumusan Masalah**

Adapun rumusan masalah dalam penelitian ini adalah sebagai berikut.

- 1 Bagaimana mengurangi *miss calculate* pada penentuan *prefiks & sufiks* dalam suatu kata
- 2 Menemukan kekurangan pada penelitian sebelumnya yang membuat minimnya akurasi nilai *Gold Standard Assessment*

### **1.3 Tujuan Penelitian**

Berdasarkan rumusan masalah di atas, adapun tujuan dari penelitian ini adalah memperbaiki nilai akurasi dari penelitian sebelumnya dengan menambahkan beberapa rule pada *source code* agar dapat meningkatnya nilai *Gold Standard Assessment*

### **1.4 Batasan Masalah**

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini berfokus pada bahasa Lampung Dialek Api.
2. Kata dalam kamus sebanyak 7627 kata dasar sebagai acuan.
3. Kata uji yang digunakan berjumlah 210 kata berimbuhan yang digunakan sebagai kata uji.

### **1.5 Manfaat Penelitian**

Manfaat dari penelitian ini adalah salah satu *tools text preprocessing* bahasa Lampung, dan salah bentuk pelestarian bahasa, yaitu bahasa Lampung.