

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Bahasa menurut (Departemen Pendidikan Nasional; Pusat Bahasa (Indonesia), 2008) adalah sistem lambang bunyi berartikulasi yang bersifat sewenang-wenang dan konvensional yang dipakai sebagai alat komunikasi untuk melahirkan perasaan dan pikiran. Bahasa juga bisa diartikan perkataan-perkataan yang dipakai oleh suatu bangsa (suku bangsa, negara, daerah, dan sebagainya). Negara Indonesia adalah salah satu negara multikultur terbesar di dunia, hal ini dapat terlihat dari kondisi sosiokultural maupun geografis Indonesia yang begitu kompleks, beragam, dan luas. “Indonesia terdiri atas sejumlah besar kelompok etnis, budaya, agama, dan lain-lain yang masing-masing plural (jamak) dan sekaligus juga heterogen “aneka ragam” (Kusumohamidjojo, 2000:45)”. Salah satunya adalah sebuah provinsi paling selatan di Pulau Sumatra yaitu Lampung, dengan ibu kota Bandar Lampung. Provinsi ini memiliki dua kota besar yaitu Kota Bandar Lampung dan Kota Metro serta 13 kabupaten, Bahasa asli dari provinsi Lampung adalah Bahasa Lampung.

Menurut Nasution, dkk (2008) Bahasa Lampung adalah bahasa daerah dan sebagai bahasa ibu bagi masyarakat di Provinsi Lampung. Bahasa Lampung dibagi menjadi 2 yaitu Pepadun dan Saibatin. Perbedaan Bahasa Lampung terletak pada perbedaan geografis. Bahasa Lampung dengan Dialek Nyow (Pepadun) adalah bahasa yang dipergunakan oleh masyarakat Lampung di wilayah nonpesisir. Adapun Bahasa Lampung Dialek Api (Saibatin) adalah bahasa yang digunakan oleh masyarakat pesisir. Sedangkan masyarakat yang tinggal di Kota Bandar Lampung ialah masyarakat yang multikultural. Karena bahasa akan berubah dan berganti

seiring berjalannya waktu maka dari itu Bahasa sangat perlu untuk dilestarikan untuk generasi muda, salah satunya adalah dengan pembuatan *Translation Engine* melihat dewasa ini telah banyak perkembangan sistem penerjemahan Bahasa yang mampu menerjemahkan Bahasa satu ke dalam Bahasa lainnya namun masih sangat sedikit sekali sistem penerjemahan yang menerjemahkan Bahasa ke Bahasa yang lebih spesifik. Sehingga sebagai bagian dari salah satu penelitian dosen di Universitas Teknokrat Indonesia untuk membuat sebuah mesin penerjemah Bahasa Indonesia ke Bahasa Lampung dan sebaliknya, penulis ikut mengambil bagian dalam penelitian tersebut untuk membuat sebuah *engine* pendeteksi bahasa, yang merupakan bagian paling awal yaitu bagian *pre-processing* dalam *engine* besar yang akan dikembangkan dan lingkupnya hanya untuk mendeteksi sebuah teks apakah itu bahasa lampung dialek A atau Bahasa Indonesia. Pendeteksian bahasa itu sendiri merupakan sebuah tugas untuk menentukan Bahasa alami didalam dokumen atau bagiannya untuk mengenali teks dalam Bahasa tertentu yang sering muncul secara alami kepada pembaca manusia yang sudah akrab dengan Bahasa tersebut.

Metode yang akan digunakan pada penelitian ini sendiri menggunakan N-Gram dan Naïve Bayes dan akan dikembangkan dengan menggunakan *Unity Engine* yang merupakan sebuah *Game Engine*, N-Gram sendiri dipilih karena dalam beberapa paper serta jurnal yang penulis telah teliti N-Gram mempunyai probabilitas yang baik dalam akurasi pendeteksian bahasa, sedangkan untuk Naïve Bayes sendiri digunakan untuk membantu penulis dalam melatih data training yang nantinya akan menjadi bagian vital pada penelitian ini, Naïve Bayes sendiri dikenal sangat buruk dalam memperkirakan sehingga menggabungkan Naïve Bayes dengan

N-Gram mungkin akan berdampak positif dilihat dari karakteristik N-Gram yang memiliki akurasi tinggi dalam memprediksi sebuah sequence. *Unity Engine* sendiri penulis pilih karena merupakan tools utama yang penulis sering pakai serta rasa penasaran penulis untuk membuat sebuah sistem atau *engine* baru menggunakan tools ini.

## 1.2 Rumusan Masalah

Dari latar belakang yang telah dijabarkan di atas, maka permasalahan yang akan diselesaikan pada penelitian ini adalah:

Bagaimana cara kerja dari algoritma N-Gram bila digabungkan dengan Naïve Bayes untuk mengklasifikasi dan mendeteksi sebuah teks yang dibuat pada sebuah *Game Engine*?

## 1.3 Batasan Masalah

Sesuai dengan judul yang tertera di atas, penulis pun membatasi pembahasan penelitian ini agar tidak menyimpang dari objek pembahasan yang diinginkan, Adapun batasan masalah pada penelitian ini adalah sebagai berikut :

1. Penelitian ini hanya dalam lingkup untuk membuat sebuah *engine* yang mampu mendeteksi sebuah inputan teks kedalam Bahasa Lampung dialek A atau Bahasa Indonesia.
2. *Engine* hanya akan dibuat pada *Unity Engine* dengan *User Interface* sederhana.

#### **1.4 Tujuan Penelitian**

Penelitian ini bertujuan untuk menghasilkan sebuah *Engine* dengan metode N-Gram dan Naïve bayes classifier yang mampu mendeteksi sebuah teks apakah berbahasa Lampung dialek A atau Bahasa Indonesia dan sebaliknya menggunakan *Unity Engine*.

#### **1.5 Manfaat atau Kontribusi Penelitian**

Manfaat dan Kontribusi dari Penelitian ini mampu membantu melestarikan Bahasa Lampung yang kurang populer belakang ini, dengan adanya kamus Bahasa Lampung yang mampu mendeteksi sebuah Bahasa yang menjadi salah satu bagian dari sebuah penelitian kamus Bahasa Lampung oleh dosen Universitas Teknokrat Indonesia.