

## 1. PENDAHULUAN

### Latar Belakang

Bahasa mempunyai peranan penting dalam kehidupan sehari-hari, karena Bahasa merupakan alat komunikasi antar manusia dengan manusia lainnya. Terdapat banyak sekali ragam bahasa yang ada di dunia, termasuk di Indonesia. Indonesia merupakan negara yang mempunyai keragaman bahasa serta budaya, dimana terdapat 726 bahasa daerah dengan 719 bahasa yang masih aktif dipergunakan sehari-hari di seluruh penjuru Indonesia [1]. Salah satunya ialah Bahasa Sunda dan Bahasa Jawa. Bahasa Sunda merupakan salah satu bahasa daerah yang digunakan oleh penduduk suku sunda, seperti di Jawa Barat dan Banten untuk saling berkomunikasi satu sama lain. Sedangkan Bahasa Jawa merupakan bahasa daerah yang digunakan oleh penduduk suku Jawa, seperti Jawa Tengah, Yogyakarta, dan Jawa Timur.

Menurut data dari Ethnologue bahasa Jawa merupakan bahasa paling banyak penuturnya di Indonesia dengan jumlah penutur mencapai kurang lebih 100.000.000 jiwa. Sedangkan Bahasa Sunda menempati peringkat dua di Indonesia dengan jumlah penutur kurang lebih 42.000.000 jiwa [2]. Sebagai masyarakat Indonesia, sudah sewajibnya kita melestarikan dan mengenalkan Bahasa daerah. Salah satu cara melestarikan Bahasa daerah ialah dengan mesin penerjemah [3].

Mesin Penerjemah otomatis (MP) adalah alat yang dapat menerjemahkan secara otomatis dari satu bahasa (sumber) ke bahasa lain (target). Terdapat beberapa metode dalam melakukan penerjemahan, salah satunya ialah mesin penerjemah statistik. Mesin Penerjemah Statistik merupakan salah satu jenis alat terjemahan di mana hasil terjemahan dihasilkan berdasarkan model statistik yang parameternya diambil dari hasil analisis korpus paralel [4]. Penerapan mesin penerjemah statistik memiliki beberapa keuntungan salah satunya adalah durasi pengembangan yang lebih cepat. Selain itu dilihat dari sisi keakuratan, pendekatan statistik memiliki hasil keakuratan yang lebih baik dibandingkan dengan pendekatan berbasis aturan [5].

Dalam penelitian ini, telah dibangun suatu Mesin penerjemah berbasis statistik dimana pengujian dilakukan dengan memasukan bahasa Sunda dan akan diterjemahkan ke dalam bahasa Jawa. Diharapkan hasil terjemahan memiliki akurasi yang cukup baik. Dalam penelitian ini, Mesin penerjemah hanya bisa menerjemahkan satu arah saja yaitu Bahasa Sunda-Jawa, dimana Bahasa Sunda merupakan Bahasa sumber, dan Bahasa Jawa sebagai Bahasa target. Adapun jumlah korpus yang digunakan adalah 1233 baris korpus paralel dan 1273 korpus monolingual.

### Topik dan Batasannya

Bedasarkan pemaparan pada latar belakang, saat ini terdapat mesin penerjemah statistik yang dapat menerjemahkan Bahasa sunda ke Bahasa jawa, yaitu Google Translate, namun hasil akurasi dari Google Translate sendiri belum cukup akurat, maka dari itu Rumusan masalah dalam Tugas Akhir ini adalah Implementasi Sistem Mesin Penerjemah Berbasis Statistik bahasa Sunda- Jawa, dan mencari tau berapa nilai akurasi dari hasil mesin penerjemah statistik Bahasa Sunda-Jawa.

Dalam penelitian ini, Mesin penerjemah hanya bisa menerjemahkan satu arah saja yaitu Bahasa Sunda-Jawa, dimana Bahasa Sunda merupakan Bahasa sumber, dan Bahasa Jawa sebagai Bahasa target. Adapun jumlah korpus yang digunakan adalah 1233 baris korpus paralel dan 1273 korpus monolingual.

### Tujuan

Tujuan dari Tugas Akhir adalah Membangun Mesin penerjemah otomatis berbasis statistik Bahasa Sunda - Bahasa Jawa dan menghasilkan mesin penerjemah statistik Bahasa Sunda-Jawa dengan hasil akurasi yang cukup baik.

### Organisasi Tulisan

1) Studi terkait

Pada tahap ini dilakukan penjabaran teori yang mendukung dan berkaitan dengan Penelitian ini,

2) Sistem yang dibangun

Pada tahap ini dibangun mesin penerjemah statistic berbasis frasa yang telah melalui proses *preprocessing*, *training*, *testing* dan evaluasi.

3) Evaluasi

Pada tahap ini dilakukan penjabaran dari hasil pengujian system yang telah dibangun.

4) Kesimpulan.

Pada tahap ini menyimpulkan seluruh proses dan hasil yang diperoleh saat penelitian.