

1. PENDAHULUAN

1.1 Latar belakang Masalah

Bahasa Indonesia adalah alat yang mampu menjembatani penduduk Indonesia yang terdiri dari berbagai suku dan bahasa untuk dapat berkomunikasi satu sama lainnya. Dalam Bahasa Indonesia dikenal beberapa jenis kata dalam sebuah kalimat seperti kata kerja, kata sifat, kata keterangan, subjek dan lain sebagainya. Setiap jenis kata tersebut tentunya memiliki fungsi yang berbeda-beda. Dalam menentukan jenis kata tersebut secara manual tentunya diperlukan waktu dan biaya yang tidak sedikit. Oleh karena itu, diperlukan suatu cara untuk menentukan jenis kata secara otomatis dengan teknik yang dinamakan *Part-Of-Speech Tagging*.

Part-Of-Speech Tagging atau yang biasa disingkat dengan POS Tagging merupakan salah satu teknik dalam *Data Mining* untuk mengenali jenis-jenis kata yang ada. POS Tagging ini sangat berguna dalam berbagai proses dalam pemrosesan bahasa Natural (NLP) seperti pengolahan teks suara, teks dan ilmu bahasa. Ada beberapa metode yang dapat digunakan untuk melakukan Tagging ini, salah satu diantaranya adalah *Hidden Markov Model*, Ruled based, Conditional Random Field dan lainnya.

POS Tagging dengan menggunakan metode *Hidden Markov Model* memiliki beberapa kekurangan, salah satunya adalah penanganan OOV atau Out-of-Vocabulary. Penanganan OOV ini adalah penanganan terhadap kata yang tidak diketahui atau kata yang sebelumnya belum pernah muncul pada saat training, sehingga hal ini dapat mengurangi akurasi dari proses tagging tersebut karena kemungkinan pemberian tag yang salah cukup tinggi. Untuk itu perlu dilakukan sebuah proses yang dapat digunakan untuk menangani masalah seperti ini.

Dalam Bahasa Indonesia sendiri OOV dapat disebabkan oleh beberapa faktor salah satunya adalah perubahan jenis kata akibat adanya penambahan imbuhan terhadap sebuah kata, misalnya kata “makan” merupakan kata kerja mendapatkan imbuhan berupa akhiran “an” menjadi “makanan” yang termasuk kata benda. Akan tetapi penanganan tersebut masih terdapat kekurangan yaitu bagaimana jika OOV tersebut merupakan sebuah kata dasar dan bukan kata berimbuhan tentu saja akan mengalami kesulitan. Tugas akhir ini akan menangani masalah OOV pada POS Tagging Hidden Markov Model dengan memanfaatkan karakteristik algoritma *Forward* dan *Backward* yang ada pada metode tersebut.

1.2 Rumusan Masalah

Penelitian ini dilakukan untuk menganalisis performansi metode Supervised Hidden Markov Model Bigram dalam menentukan jenis kata berbahasa Indonesia dengan penanganan oov menggunakan algoritma *Forward* dan *Backward* dan menarik kesimpulan dari hasil analisis.

1.3 Tujuan

Tujuan dari penelitian ini adalah :

1. Membuat aplikasi penentuan jabatan kata dalam kalimat berbahasa Indonesia.
2. Menganalisis performansi dari Pos Tagging menggunakan Hidden Markov Model dengan penanganan OOV dalam menentukan jabatan kata dalam kalimat berbahasa Indonesia.

1.4 Manfaat

Manfaat dari tugas akhir ini adalah:

Bagi penyusun :

1. Dapat memahami penanganan OOV pada *Hidden Markov Model Pos Tagging*

Bagi dunia pendidikan :

1. Dapat menjadi perbandingan atau referensi dalam pembuatan kamus data berbahasa Indonesia untuk pengolahan data.
2. Dapat menambah pengetahuan mengenai pemrosesan bahasa natural.

1.5 Batasan Masalah

Berikut ini merupakan batasan masalah dari penelitian:

1. Penelitian ini dilakukan pada *Pos Tagging Hidden Markov Model* bigram.
2. *Input* berupa kalimat dalam berita berbahasa Indonesia. Jumlah kalimat yang diinputkan adalah sebuah kalimat dan sebuah paragraf.
3. *Output* berupa kata-kata beserta jabatan katanya dalam kalimat yang diinputkan.

1.6 Hipotesa

Algoritma *Forward* dan *Backward* dapat menangani masalah Out-Of-Vocabulary yang terjadi pada metode HMM.

1.7 Metodologi Penyelesaian Masalah

1. Studi literatur

Pada tahap ini penulis mencari referensi-referensi yang berkaitan dengan *POS Tagging*, metode *Hidden Markov Model* dan algoritma *Backward* dan sumber lain yang masih relevan dengan judul Tugas Akhir

2. Pengumpulan data

Mencari data berupa daftar tag dan kata dalam Bahasa Indonesia sebagai bahan dalam pembuatan aplikasi.

3. Perancangan

Dalam tahap perancangan ini, akan dilakukan beberapa kegiatan seperti :

- a. Perancangan desain antarmuka dari aplikasi yang dibangun.
- b. Merancang penerapan Metode Hidden Markov Model dalam skema yang telah dibuat.

4. Implementasi

Tahap implementasi dari sistem ini adalah dengan membangun sebuah sistem yang telah dirancang untuk penentuan jenis kata Bahasa Indonesia menggunakan metode Supervised Hidden Markov Model dan algoritma *Backward* untuk penanganan OOV. Dalam mengimplementasikannya dibutuhkan perangkat keras dan perangkat lunak yang memadai. Berikut adalah daftar perangkat keras dan perangkat lunak yang diperlukan, yaitu :

a. Spesifikasi perangkat keras

1. Processor : Intel(R) Core(TM) i5-2450M CPU @2.50GHz (4 CPUs)
2. Memory : 4,00 GB
3. Harddisk : 500 GB

b. Spesifikasi perangkat lunak

1. Sistem Operasi : Microsoft Windows Seven
2. Bahasa Pemograman : Java

5. Pengujian

Pada tahap pengujian akan dilakukan 2 (dua) buah proses pengujian secara berulang kali terhadap aplikasi yang dibangun yaitu :

- a. Pengujian apakah aplikasi yang dibangun sudah siap dan tidak terjadi kesalahan.
- b. Pengujian dengan menggunakan data sehingga menghasilkan keluaran berupa dokumen dengan kalimat ber-tag.

6. Analisis hasil

Hasil yang didapat kemudian dari pengujian dengan menggunakan aplikasi akan dianalisis, dilakukan serangkaian pengamatan sehingga diperoleh suatu kesimpulan. Adapun hal yang akan di analisis adalah bagaimana kesesuaian output yang dihasilkan, serta mengukur performansi dari algoritma yang dipakai.

7. Pembuatan laporan

Keseluruhan data, aplikasi, analisis hasil dll, disertakan pada pembuatan laporan tugas akhir. Penyusunan laporan dilakukan secara bertahap sampai proses pembuatan aplikasi selesai.

1.8 Sistematika Penulisan

Penelitian ini diuraikan dengan sistematika sebagai berikut :

Bab I Pendahuluan

Pada bab ini berisi uraian mengenai latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi penelitian serta sistematika penulisan.

Bab II Landasan Teori

Bab ini berisi literatur yang relevan dengan permasalahan yang diangkat yang diperoleh dari berbagai sumber mengenai Tagging, Hidden Markov Model, dan algoritma *Backward* dan semua literatur yang berkaitan dengan Tugas Akhir ini.

Bab III Analisis dan Perancangan Sistem

Bab ini berisi uraian mengenai perangkat lunak seperti spesifikasi kebutuhan perangkat, perancangan struktur data dan spesifikasi proses dari perangkat lunak yang dibuat.

Bab IV Implementasi dan Analisis Pengujian Hasil

Bab ini berisi uraian mengenai pengolahan data klasifikasi yang digunakan serta analisis hasil pengujian perangkat lunak.

Bab V Kesimpulan dan Saran

Bab ini berisi kesimpulan dari seluruh sistem yang dibuat serta saran untuk pengembangan perangkat lunak.