

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Part of Speech Tag atau yang sering disebut POS Tag bertujuan untuk mengklasifikasikan setiap kata ke dalam kelas kata yang sesuai (kata kerja, kata sifat, dll). POS Tag berperan terhadap bidang NLP lainnya seperti *Question-Answering*, *Speech Recognition*, *information retrieval system*, *machine translation system*, *word sense disambiguation system*, dan sebagainya [1].

Penelitian POS Tag telah dilakukan pada teks bahasa Indonesia. Pada penelitian tersebut, metode yang digunakan adalah rule based dengan kelas kata berjumlah 23 buah dan memiliki tingkat akurasi sebesar 79% [2]. Berdasarkan hasil dari penelitian tersebut, Rule Based memiliki masalah terkait ambiguitas yang mengakibatkan penurunan tingkat akurasi. Hal ini disebabkan oleh sulitnya pembuatan aturan Rule Based dalam menangani struktur dan *grammar* bahasa. Masalah tersebut dapat ditangani dengan *machine learning*. Ada beberapa metode yang belum diterapkan pada POS Tag bahasa Indonesia diantaranya adalah SVM, KNN, dan Decision Tree. Dalam penelitian mengenai komparasi metode pada POS Tag bahasa thailand, metode yang paling baik dari sisi akurasinya adalah SVM diikuti oleh Decision Tree dan yang terakhir yaitu KNN [3]. Penelitian lainnya yang menunjukkan SVM memiliki akurasi yang lebih baik dalam membandingkan ketiga metode tersebut adalah penelitian yang dilakukan oleh Amr [4].

Penelitian mengenai penanganan ambiguitas telah dilakukan dengan menggunakan SVM dapat kita lihat pada penelitian yang dilakukan oleh Gimenez untuk POS Tag bahasa Inggris dan bahasa Spanyol. Nilai akurasi yang didapat sangatlah baik yaitu sebesar 97,16% untuk bahasa Inggris dan 98,86% untuk bahasa Spanyol [5] [6]. Selain penelitian yang dilakukan oleh Gimenez, SVM pada POS Tag bahasa Inggris juga telah dilakukan oleh Nakagawa. Data yang digunakan sebanyak 2.416 kalimat atau 56.684 kata. Penelitian itu menghasilkan akurasi sebesar 97,11% [7].

Berdasarkan penelitian-penelitian tersebut, POS Tag bahasa Indonesia yang akan dibuat menggunakan SVM kemungkinan memiliki akurasi yang tinggi juga. Dengan alasan lainnya juga bahwa metode SVM belum pernah diterapkan pada POS Tag bahasa Indonesia maka diputuskan untuk membuat penelitian mengenai sistem POS Tagger bahasa Indonesia menggunakan SVM.

1.2 Identifikasi Masalah

Berdasarkan latar belakang di atas maka dapat diidentifikasi bahwa masalah yang dihadapi pada penelitian sebelumnya adalah sulitnya mengatasi ambiguitas karena beragamnya struktur dan *grammar* suatu bahasa yang menyebabkan menurunnya akurasi.

1.3 Maksud dan Tujuan

Penelitian ini bermaksud untuk membangun sebuah sistem POS Tag bahasa Indonesia menggunakan metode SVM.

Adapun tujuan penelitian ini yaitu mengukur akurasi yang dihasilkan oleh SVM pada POS Tag bahasa Indonesia.

1.4 Batasan Masalah

Batasan masalah adalah upaya untuk membatasi ruang lingkup masalah pada penelitian ini. Batasan-batasan pada penelitian ini, yaitu:

- a. Data masukan
 1. Teks masukan yang digunakan adalah corpus “Indonesian_Manually_Tagged_Corpus” dengan format file berupa *Tab Separated Values*(.TSV) sebagai data *training* [8].
 2. Teks masukan dari PAN Localization dengan format file berupa txt sebagai data *testing*.
- b. Proses
 1. Menggunakan metode SVM dengan kernel Linear karena efisiensinya yang baik dan memiliki akurasi yang tinggi [5] .
 2. Memakai teknik *multiclass one against all* [9] yang lebih baik dibandingkan *one against one*.

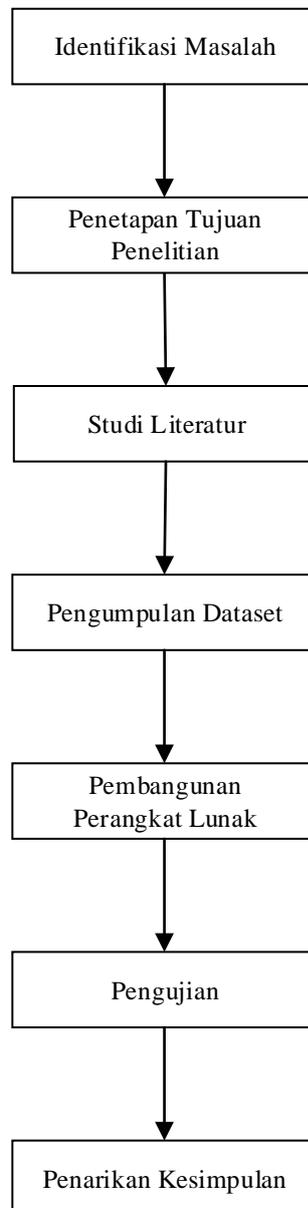
3. Pada ekstraksi fitur, digunakan 14 fitur diantaranya Caps, In_Cap, All_Cap, All_low, Hyp, Num, Prefix('me-', 'pe-', 'se-', 'ke', 'be-', 'di-'), dan Suffix('-an', '-kan') [5] [10].

c. Data Keluaran

Data yang dikeluarkan berupa daftar kata yang telah diklasifikasikan berdasarkan kelas katanya.

1.5 Metode Penelitian

Dalam mendukung kelancaran penelitian ini, metode penelitian menggunakan metode Deskriptif [11]. Metode ini digunakan karena penelitian yang dilakukan berdasarkan fakta, bukan opini dan menggunakan perhitungan matematik. Alur metode ini bisa dilihat pada Gambar 1.1.



Gambar 1.1 Alur Metode Penelitian Deskriptif

1.5.1 Identifikasi Masalah

Identifikasi masalah adalah proses memilih dan merumuskan masalah yang akan diteliti. Masalah tersebut harus dapat diselidiki untuk diselesaikan dengan sumber yang ada.

1.5.2 Penetapan Tujuan Penelitian

Menentukan tujuan dari penelitian yang didasarkan pada rumusan masalah. Rumusan masalah didapat dari proses identifikasi masalah.

1.5.3 Studi literatur

Studi literatur ini didapat dari berbagai sumber diantaranya buku, jurnal, serta sumber ilmiah lainnya yang berkaitan dengan *Part of Speech Tag* dan metode *Support Vector Machine*.

1.5.4 Pengumpulan Dataset

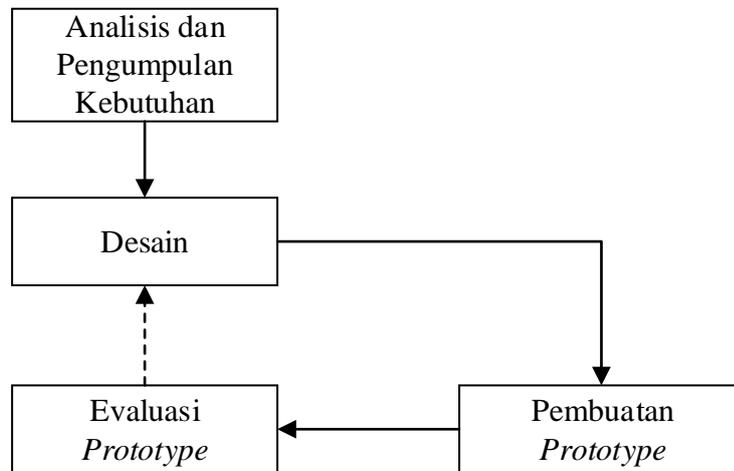
Dalam melakukan penelitian ini dibutuhkan dataset baik itu untuk proses *training* maupun proses *testing*. Dataset yang akan digunakan didapat dari hasil penelitian yang dilakukan oleh Dinakaramani yang mana dataset tersebut berupa teks bahasa Indonesia dengan 23 kelas kata sebagai dataset *training* [8]. Sedangkan dataset *testing* diambil dari PAN Localization.

1.5.5 Pembangunan Perangkat Lunak

Perangkat lunak dibangun karena penelitian ini memiliki maksud untuk membuat sistem POS Tagger bahasa Indonesia menggunakan metode SVM. Metode pembangunan perangkat lunak pada penelitian ini akan menggunakan model *prototype*. Blok diagram model *prototype* dapat dilihat pada Gambar 1.2 [12]. Model ini digunakan bertujuan untuk mendapat akurasi yang baik di mana apabila masih ada yang kurang dari sistem *prototype* maka sistem akan diperbaiki. Alur dimulai dari analisis dan pengumpulan kebutuhan hingga evaluasi *prototype*, apabila sistem belum sesuai dengan kebutuhan penelitian maka sistem akan diperbaiki dan alurnya akan diulang dari proses desain menuju proses membangun *prototype*, dan dievaluasi kembali.

1.5.6 Pengujian

Hasil pengujian pada penelitian ini adalah nilai akurasi dari sistem dengan menggunakan metode *Support Vector Machine* pada *Part of Speech Tag* bahasa Indonesia. Hal ini dimaksudkan untuk menguji tercapainya tujuan dari penelitian ini yaitu mendapatkan akurasi.



Gambar 1.2 Alur Metode *Prototype*

Berikut penjelasan tentang tahapan pada alur metode *prototype*.

1. Analisis dan Pengumpulan, mengidentifikasi kebutuhan dan proses yang diperlukan dalam membuat sistem POS Tag-SVM. Selain itu, pada tahap ini juga dilakukan proses mengumpulkan kebutuhan tersebut.
2. Desain, merupakan tahap dalam merancang antarmuka dan alur sistem dari sudut pengguna.
3. Pembuatan prototype, tahapan ini merupakan pembuatan kode program yang akan menghasilkan *prototype* sistem.
4. Evaluasi *Prototype*, menguji dan menyesuaikan sistem dengan kebutuhan sistem yang akan dikembangkan. Pada tahap ini akan dicari tingkat akurasi dan menguji fungsional sistemnya.

1.5.7 Penarikan Kesimpulan

Berdasarkan hasil dari pengujian dapat disimpulkan bahwa metode tersebut baik atau tidak diterapkan pada *Part of Speech Tag* Bahasa Indonesia. kesimpulan tersebut merupakan hasil dari penelitian ini.

1.6 Sistematika Penulisan

Sistematika penulisan laporan penelitian ini disusun untuk memberikan gambaran umum tentang kegiatan penelitian yang akan dilakukan. Sistematika penulisan laporan penelitian ini adalah :

BAB 1 PENDAHULUAN

Bab ini berisi penjelasan tentang penelitian yang sedang dilakukan. Bab ini meliputi latar belakang masalah, identifikasi masalah, maksud dan tujuan, batasan masalah, metode penelitian, dan sistematika penulisan.

BAB 2 TINJAUAN PUSTAKA

Bab ini berisi tentang teori-teori yang digunakan dalam mendukung penelitian Penerapan SVM pada POS Tag Bahasa Indonesia ini. Teori-teori itu meliputi *Part Of Speech Tag*, *Support Vector Machine*, Python, dan Sklearn.

BAB 3 ANALISIS DAN PERANCANGAN SISTEM

Bab ini menjelaskan tentang tahapan dalam pembuatan sistem Penerapan SVM pada POS Tag Bahasa Indonesia. Tahapan tersebut diantaranya analisis masalah, analisis sistem, analisis evaluasi sistem, fungsional, nonfungsional, diagram konteks, DFD, dan perancangan antarmuka.

BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM

Bab ini menjelaskan implementasi yang dilakukan pada sistem dengan menjalankan sistem tersebut dari mulai memasukan data, *training*, *testing*, hingga didapat akurasi.

BAB 5 KESIMPULAN DAN SARAN

Bab ini menjelaskan hasil dari penelitian yang sudah dikerjakan mengenai Penerapan Metode *Support Vector Machine* pada *Part of Speech Tag* Bahasa Indonesia. Selain itu juga terdapat hal-hal yang penulis dapat simpulkan dan sarankan untuk digunakan dalam pengembangan penelitian selanjutnya.

