

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Pengenalan entitas bernama atau *Named Entity Recognition* (NER) adalah sebuah topik keilmuan yang membahas pemrosesan bahasa alami dengan mengekstraksi informasi pada sebuah teks dengan memberikan entitas pada setiap kata yang dianggap penting dan menentukan entitasnya ke dalam kategori yang sudah didefinisikan seperti nama orang, lokasi, organisasi, waktu, dan kuantitas dalam suatu teks. NER juga merupakan komponen penting yang mendasari banyak aplikasi *Natural Language Processing* (NLP) seperti *information extraction* (IE), *Question Answering*, *Text Summarization*, *Information Retrieval*, dan lain-lain.

Beberapa metode *machine learning* yang sudah digunakan dalam NER diantaranya menggunakan metode CRF [1] dengan akurasi sebesar 90%, LSTM-CRF [2] dengan nilai F1-score sebesar 73%, Naive Bayes [3] dengan F1-score sebesar 81%, dan SVM untuk teks Bahasa Indonesia menghasilkan F1-score sebesar 50.4% [4]. Pada penelitian [4] NER pada teks bahasa Indonesia mengandalkan hasil dari penggunaan ekstraksi fitur untuk melakukan klasifikasi sehingga performa yang dihasilkan masih rendah. Namun pada penelitian lain metode SVM menghasilkan performa yang cukup baik dalam melakukan klasifikasi teks berita online yang menghasilkan F1-score sebesar 93.14 % [5].

Berdasarkan pemaparan diatas metode SVM menghasilkan performa yang baik dalam melakukan klasifikasi teks namun dalam kasus NER pada teks bahasa Indonesia nilai performa yang dihasilkan masih yang rendah. Oleh karena itu dalam penelitian ini data teks akan direpresentasikan dengan metode *one hot encoding* terlebih dahulu dengan tambahan ekstraksi fitur yang akan digunakan. Metode *one hot encoding* dipilih karena menghasilkan akurasi sebesar 90 % [6]. Berdasarkan pemaparan tersebut, maka dalam penelitian ini data akan direpresentasikan dengan metode *one hot encoding* dengan tambaha ekstraksi fitur yang akan digunakan

dengan harapan dapat meningkatkan performa dari metode SVM untuk pengenalan entitas bernama pada teks bahasa Indonesia.

1.2 Identifikasi Masalah

Berdasarkan pemaparan dari latar belakang di atas, maka masalah yang teridentifikasi adalah rendahnya nilai akurasi SVM pada kasus NER untuk teks Bahasa Indonesia.

1.3 Maksud dan Tujuan

Berdasarkan uraian latar belakang, maka maksud dari penelitian ini adalah untuk menerapkan *one hot encoding* pada metode *Support Vector Machine* dalam pengenalan entitas bernama pada teks Bahasa Indonesia. Tujuan dari penelitian ini adalah untuk meningkatkan performa yang dicapai berdasarkan F1-score dengan mengimplementasikan *one hot encoding* pada metode *Support Vector Machine* untuk pengenalan entitas bernama pada teks bahasa Indonesia.

1.4 Batasan Masalah

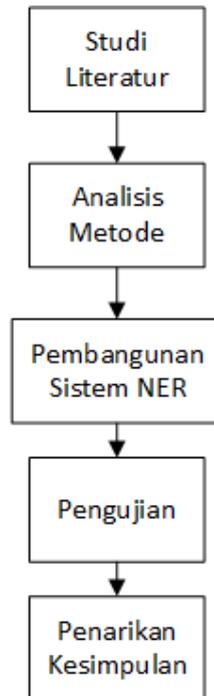
Batasan masalah dalam penelitian ini adalah sebagai berikut.

1. Masukan
 - a. Format file yang akan digunakan adalah *.txt
 - b. Data set yang akan digunakan berasal dari teks berita politik.
 - c. Data set yang digunakan berasal dari penelitian Rusliani [7].
2. Proses
 - a. Dalam penelitian digunakan enam kelas yaitu nama orang, lokasi, waktu, organisasi, kuantitas dan lainnya.
3. Keluaran
 - a. Hasil keluaran dari sistem berupa label kelas entitas.

1.5 Metode Penelitian

Metode penelitian yang akan digunakan dalam penelitian ini adalah metode kuantitatif [8]. Metode ini digunakan karena pada penelitian ini data yang

digunakan berbentuk angka dan analisis menggunakan statistik. Alur penelitian dari penelitian ini dapat dilihat pada Gambar 1.1 di bawah ini.



Gambar 1.1 Alur Penelitian

1.5.1 Studi Literatur

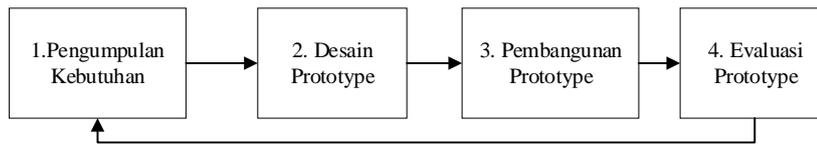
Studi literatur dilakukan mengumpulkan bahan bacaan seperti buku, jurnal, paper, situs dari internet yang ada kaitannya dengan metode SVM pada kasus NER.

1.5.2 Analisis Metode

Pada tahap ini akan dilakukan analisis terhadap metode yang akan digunakan, seperti kebutuhan data masukan yang akan diolah, alur proses dari metode *one hot encoding*, ekstraksi fitur dan metode SVM.

1.5.3 Pembangunan Sistem NER

Metode pembangunan perangkat lunak yang digunakan adalah metode *prototype* [9]. Metode ini dipilih karena pada setiap tahapnya akan dievaluasi dan dirancang kembali jika ada tahap yang tidak sesuai dengan yang diharapkan, sehingga ketika diuji akan didapatkan sistem yang sesuai yang diinginkan. Adapun tahapan metode *prototype* dapat dilihat pada Gambar 1.2.



Gambar 1.2 Metode Prototype

1. Pengumpulan Kebutuhan

Kegiatan yang dilakukan pada tahap ini adalah mengumpulkan kebutuhan perangkat keras, perangkat lunak, dataset yang diperlukan untuk membuat sistem NER.

2. Desain *Prototype*

Pada tahap ini dilakukan perancangan pembuatan perangkat lunak. Pembangunan perangkat lunak menggunakan pendekatan objek maka akan berfokus pada pembuatan UML, dan perancangan antarmuka.

3. Pembangunan *Prototype*

Pada tahap ini dilakukan pembangunan *prototype* meliputi *preprocessing*, proses *training*, dan proses *testing* dengan menggunakan metode SVM.

4. Evaluasi *Prototype*

Pada tahap ini *prototype* yang telah dibangun pada tahap sebelumnya akan dilakukan pengujian untuk melihat apakah *prototype* sudah berjalan dengan baik. Jika masih ada kekurangan, maka *prototype* akan diperbaiki dengan cara kembali ke tahapan sebelumnya.

1.5.4 Pengujian

Pengujian dilakukan dengan mengukur seberapa besar nilai akurasi dan performa yang dihasilkan dari implementasi metode one hot encoding pada SVM untuk kasus NER pada teks Bahasa Indonesia. Pengujian ini dilakukan untuk mengetahui bagaimana performa yang dihasilkan oleh sistem.

1.5.5 Penarikan Kesimpulan

Berdasarkan hasil pengujian langkah selanjutnya yaitu penarikan kesimpulan dari hasil penerapan *one hot encoding* pada metode SVM untuk pengenalan entitas bernama.

1.6 Sistematika Penulisan

Sistematika penulisan disusun untuk memberikan gambaran secara umum mengenai permasalahan dan pemecahannya. Sistematika penulisan laporan penelitian ini adalah sebagai berikut.

BAB 1 PENDAHULUAN

Bab ini membahas mengenai latar belakang, identifikasi masalah, maksud dan tujuan, batasan masalah, metode penelitian, serta sistematika penulisan yang dimaksudkan untuk memberikan penjelasan mengenai penelitian ini.

BAB 2 LANDASAN TEORI

Bab ini berisi tentang landasan teori dan metode yang digunakan pada tugas akhir ini yaitu teori mengenai *Named Entity Recognition*, metode *Support Vector Machine*.

BAB 3 ANALISIS DAN PERANCANGAN

Bab ini berisi pemaparan mengenai analisis masalah, analisis proses, analisis kebutuhan non fungsional, analisis kebutuhan fungsional, perancangan antar muka, jaringan semantik.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi tentang tahapan yang dilakukan dalam penelitian secara garis besar, analisis dari bab-bab sebelumnya dan perancangan sistem yang dilakukan. Hasil dari analisis kemudian dilakukan pengujian sistem dengan metode *blackbox* sehingga perangkat lunak yang dibangun sesuai dengan analisis dan perancangan yang telah dilakukan.

BAB 5 KESIMPULAN DAN SARAN

Bab ini menjelaskan tentang kesimpulan yang diperoleh dari hasil analisis metode SVM untuk pengenalan entitas bernama pada teks Bahasa Indonesia dan saran yang dapat dijadikan untuk pengembangan penelitian pengenalan entitas bernama selanjutnya.