

BAB 1

PENDAHULUAN

1.1. Latar Belakang Masalah

Banyak kajian yang telah diteliti dibidang *Natural Language Processing*, salah satunya adalah ekstraksi informasi. Ekstraksi informasi merupakan proses pengambilan informasi terstruktur secara otomatis. Informasi terstruktur seperti surat keputusan. Surat keputusan adalah surat yang berisi suatu keputusan yang dibuat oleh pimpinan suatu organisasi atau lembaga pemerintahan berkaitan dengan kebijakan organisasi atau lembaga tersebut [1].

Penelitian sebelumnya mengenai ekstraksi informasi pernah dilakukan oleh Arbi [2] dan Dimas [3] dengan menggunakan metode rule-based. Penelitian yang dilakukan oleh Arbi [2] mengangkat kasus ekstraksi informasi pada dokumen teks novel. Penelitian ini menggunakan metode *rule-based*, hasil dari penelitian ini menghasilkan nilai akurasi sebesar 100%. Penelitian yang dilakukan oleh Dimas [3] mengangkat kasus ekstraksi informasi pada teks skripsi. Penelitian ini juga menggunakan metode rule-based, hasil dari penelitian ini menghasilkan nilai akurasi sebesar 100%. Walaupun dalam penelitian ini menghasilkan nilai akurasi yang tinggi, namun memiliki kekurangan yaitu data yang digunakan harus memiliki format dokumen yang sama. Semakin banyak format dokumen yang berbeda, maka semakin banyak aturan yang harus dibuat. Oleh karena itu, dibutuhkan algoritma yang dapat mengklasifikasi dokumen dengan format beragam. Salah satu algoritma yang dapat menangani kasus tersebut dengan menggunakan algoritma deep learning. Hal ini didasari oleh penelitian sebelumnya yang dilakukan oleh B. Wang dan M Liu [4] membuktikan bahwa model deep learning baik dalam menyelesaikan kasus analisis sentiment berdasarkan aspek. Penelitian tersebut menggunakan algoritma CNN. Hasil dari penelitian tersebut menghasilkan akurasi sebesar 79.3%. Selanjutnya penelitian yang dilakukan oleh P. Liu dkk [5] dalam kasus yang sama dengan membandingkan arsitektur dari RNN. Pada Elman RNN menghasilkan nilai akurasi sebesar 81.36% , untuk LSTM menghasilkan akurasi sebesar 79.43% dan untuk Jordan RNN menghasilkan akurasi sebesar 78.83%. Berdasarkan uraian yang

telah dipaparkan, maka pada penelitian ini akan menerapkan metode ERNN kedalam sistem ekstraksi informasi pada surat keputusan berbahasa indonesia.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan diatas, maka dapat dirumuskan masalah yang dihadapi bahwa sistem ekstraksi informasi dengan menggunakan *rule-based* tidak cocok dengan format dokumen yang beragam, sehingga diperlukan penerapan sistem ekstraksi informasi pada dokumen yang formatnya beragam salah satunya adalah dengan menggunakan metode ERNN.

1.3. Maksud dan Tujuan Penelitian

Maksud dari penelitian ini adalah untuk melakukan analisis dan mengimplementasikan *Elman Recurrent Neural Network* dalam kasus ekstraksi informasi pada surat keputusan berbahasa indonesia. Adapun tujuan dari penelitian ini adalah sistem dapat mengetahui informasi yang ada pada surat keputusan serta mengukur seberapa besar akurasi yang dihasilkan dari metode *Elman Recurrent Neural Network* dalam kasus ekstraksi informasi pada surat keputusan.

1.4. Batasan Masalah

Agar masalah yang diteliti tidak menyimpang , lebih terarah maka dalam penelitian ini diberikan batasan-batasan sebagai berikut :

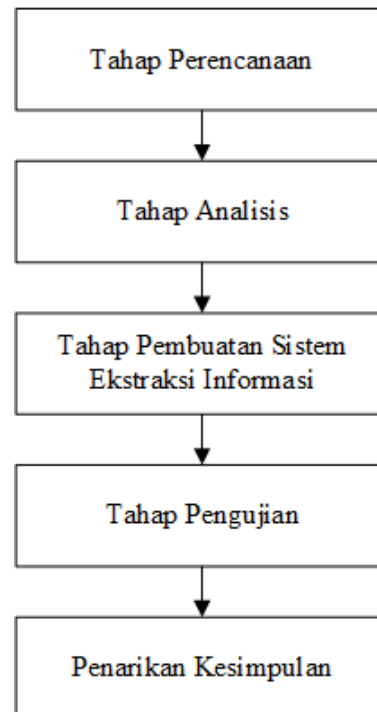
1. Dokumen yang dijadikan objek penelitian adalah dokumen surat keputusan berbahasa Indonesia.
2. Dokumen yang digunakan hanya merujuk pada satu orang. Hal ini didasari, jika lebih dari 1 orang maka kompleksitas algoritma lebih rumit.
3. Informasi yang akan diambil berupa : nomor surat, tentang, nama, yang dituju dan tanggal penetapan.

1.5. Metode Penelitian

Metode penelitian yang akan digunakan dalam penelitian ini adalah metode kuantitatif. Menurut Saifuddin Azwar [6] metode kuantitatif merupakan penelitian yang menekankan pada data-data numerikal (angka) yang diolah dengan metode statistika. Sedangkan menurut Sugiyono [7] metode penelitian kuantitatif merupakan metode penelitian yang digunakan untuk meneliti pada populasi atau sampel tertentu, teknik pengambilan sampel pada umumnya dilakukan secara

random, pengumpulan data menggunakan instrumen penelitian, analisis data bersifat kuantitatif atau statistik dengan tujuan untuk menguji hipotesis yang telah ditetapkan.

Adapun alur penelitian yang akan dilakukan ditunjukkan pada Gambar 1.1.



Gambar 1. 1 Alur Penelitian

Adapun penjelasan dari Gambar 1.1 sebagai berikut :

1.5.1. Tahap Perencanaan

Adapun pada tahap perencanaan dalam penelitian ini memiliki beberapa tahapan, yaitu :

1. Identifikasi masalah

Pada tahap pertama ialah mengidentifikasi masalah pada ekstraksi informasi surat keputusan. Pada tahap kedua, yaitu merumuskan masalah metode atau fitur-fitur yang akan digunakan pada kasus ekstraksi informasi.

2. Studi literatur

Pada tahap ini bertujuan untuk menambah pengetahuan, pemahaman mengenai algoritma *Elman Recurrent Neural Network* dan ekstraksi

informasi pada surat keputusan. Studi literature yang dilakukan meliputi mengumpulkan beberapa bahan atau materi dari e-book, internet, jurnal dan sumber-sumber yang berhubungan dengan permasalahan yang sedang diteliti.

3. Pengumpulan Data

Pengumpulan data dilakukan dengan cara mengumpulkan dokumen-dokumen surat keputusan yang akan dipakai sebagai data set. Adapun dokumen yang dijadikan dataset diperoleh dari instansi.

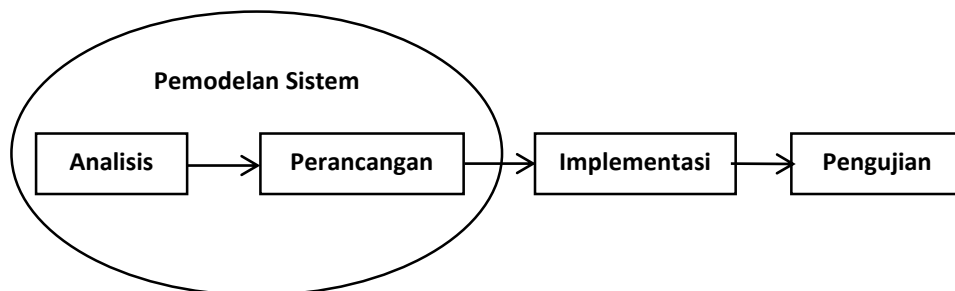
1.5.2. Tahap Analisis

Tahapan kedua dari alur penelitian ini adalah tahap analisis, adapun tahapan analisis yang dilakukan sebagai berikut :

1. Menganalisis dokumen masukan, yaitu menganalisis pola dokumen yang akan digunakan sebagai objek penelitian. Dokumen yang digunakan adalah dokumen surat keputusan yang telah di scan dan tersimpan dalam format .jpg.
2. Menganalisis metode, analisis ini dilakukan untuk mengetahui karakteristik dan performa dari algoritma *Elman Recurrent Neural Network* dalam kasus ekstraksi informasi pada surat keputusan.

1.5.3. Tahap Pembuatan Sistem Ekstraksi Informasi

Sistem yang akan dibangun adalah menggunakan model perangkat lunak *sequential linear* karena tahapan dalam penelitian ini bersifat sekuensial mulai dari pengumpulan data, tahap praproses, klasifikasi hingga pengujian yang dilakukan secara sekuensial. Model *sequential linear* [8] ditujukan pada Gambar 1.2.



Gambar 1. 2 Sequential Linear Model [8]

Adapun penjelasan dari Gambar 1.2 pada masing-masing tahap sebagai berikut :

1. Analisis kebutuhan

Pada tahap ini dilakukan analisis kebutuhan sistem yang diperlukan dalam membangun sistem ekstraksi informasi menggunakan ERNN. Analisis dilakukan terhadap data masukan, tahapan proses yang harus dilakukan, perangkat lunak, dan perangkat keras yang dibutuhkan.

2. Desain

Tahap ini merupakan penjabaran dari proses sebelumnya, pada tahap ini ditentukan seperti apa sistem yang nantinya akan dibangun, metode yang digunakan dalam proses ekstraksi informasi.

3. Implementasi

Setelah tahap desain sistem selesai, maka tahap selanjutnya adalah implementasi ke dalam bahasa pemrograman sesuai dengan desain sistem yang sebelumnya sudah dibuat.

4. Pengujian

Setelah tahap implementasi selanjutnya masuk ke tahap pengujian. Pada tahap ini akan dilakukan pengujian dari sistem yang telah dibangun. Pengujian berupa pengujian fungsionalitas dengan menggunakan metode *black box*.

1.5.4. Tahap Pengujian

Dalam tahap ini akan dilakukan pengujian dari pengimplementasian *Elman Recurrent Neural Network* kedalam sistem ekstraksi informasi, pengujian dilakukan dengan cara mengecek apakah informasi yang diperoleh dari algoritma ERNN sesuai atau tidak dari dokumen surat keputusan, kemudian informasi yang sesuai akan dihitung jumlah benarnya agar dapat dilakukan perhitungan seberapa besar akurasi yang didapat terhadap sistem ekstraksi informasi. Perhitungan akurasi menggunakan metode *confusion matrix*.

1.5.5. Penarikan Kesimpulan

Pada tahapan ini akan dilakukan penarikan kesimpulan berdasarkan hasil dari tahap pengujian dalam pengimplementasian *Elman Recurrent Neural Network* dalam kasus ekstraksi informasi pada surat keputusan.

1.6. Sistematika Penulisan

Sistematika penulisan penelitian ini disusun untuk memberikan gambaran umum mengenai penelitian yang dikerjakan. Sistematika penulisan penelitian adalah sebagai berikut :

BAB 1 PENDAHULUAN

Pada bab ini berisi penjelasan mengenai latar belakang masalah, identifikasi masalah, maksud dan tujuan, batasan masalah, metode penelitian dan sistematika penulisan yang dimaksud agar dapat memberikan gambaran dalam penelitian.

BAB 2 TINJAUAN PUSTAKA

Pada bab ini berisi teori-teori atau tinjauan dari literatur pendukung yang akan digunakan untuk mengekstrak informasi pada surat keputusan. Adapun teori-teori yang akan ditulis pada bab ini yaitu teori *text mining*, ekstraksi informasi, *preprocessing* metode *Recurrent Neural Network* serta bahasa pemrograman yang akan digunakan dalam membangun sistem.

BAB 3 ANALISIS DAN PERANCANGAN

Pada bab ini berisi penjelasan secara rinci mulai dari analisis masalah, analisis sistem, analisis tahap *preprocessing*, analisis algoritma *Recurrent Neural Network*.. Selain itu juga dijelaskan mengenai analisis kebutuhan fungsional, analisis kebutuhan non-fungsional serta perancangan antar muka.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Pada bab ini berisi mengenai hasil dari implementasi dari analisis yang telah dilakukan di BAB 3 yaitu berupa perancangan sistem yang meliputi : implementasi perangkat keras, implementasi perangkat lunak dan implementasi perangkat lunak. Pada proses pengujian, pengujian yang dilakukan berupa pengujian sistem yang meliputi pengujian fungsionalitas dengan menggunakan metode black box serta nilai akurasi terhadap sistem ekstraksi informasi.

BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dan saran dari hasil penelitian mengenai performa *Elman Recurrent Neural Network* dalam kasus ekstraksi informasi pada surat keputusan agar dapat digunakan sebagai referensi untuk penelitian selanjutnya.

