

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Ekstraksi informasi adalah sebuah proses untuk menemukan informasi terstruktur dari dokumen tidak terstruktur atau semi terstruktur[1]. Ekstraksi informasi dilakukan dengan cara mengambil semua data berdasarkan aturan yang telah dibuat, lalu dari data yang diambil dipisahkan lagi sehingga didapatkan informasi yang diinginkan. Terdapat berbagai macam bentuk informasi, seperti koran, novel, karya ilmiah, buku, jurnal, juga surat baik surat masuk, surat perintah, surat keputusan, dan sebagainya.

Dalam *Natural Language Processing (NLP)*, *Convolutional Neural Network (CNN)* terbukti efisien untuk menangkap maksud dari representasi kalimat, seperti klasifikasi, pemodelan bahasa, analisis sentimen, *tagging*, pemberian label, bahkan ekstraksi informasi. CNN biasanya diterapkan untuk pengenalan gambar, namun dapat juga digunakan sebagai *text processing*.

CNN adalah kategori *Neural Network* yang menggunakan multilayer variasi persepsi yang dirancang untuk preprocessing minimal[2]. CNN telah terbukti sangat efektif di berbagai bidang seperti pengenalan gambar atau klasifikasi. Namun, saat ini, ada banyak studi CNN yang menggunakan teks sebagai data yang diproses.

CNN menjadi salah satu algoritma yang memiliki akurasi tinggi pada klasifikasi teks, terdapat beberapa penelitian sebelumnya yang termasuk ke dalam *text processing* menggunakan CNN. Bagas Prakoso Putra[3] dengan penelitiannya tentang analisis sentimen menghasilkan akurasi sebesar 99,8% dengan rata-rata precision dan recall masing-masing sebesar 99,46% dan 97,99%. Penelitian yang dilakukan Fajar Ratnawati dan Edi Winarko[4] tentang analisis sentiment menggunakan Dynamic CNN menghasilkan nilai akurasi yang lebih tinggi dari pada Naive Bayes, yaitu 80,99%. Zhiquan Wang dan Zhiyi Qu[5] tentang klasifikasi teks dengan mengkomparasikan algoritma KNN, SVM dan CNN. Pada penelitian tersebut didapat CNN lebih unggul dengan menghasilkan nilai presisi

88,4%, recall 87,4%, F-Measure 87,9%, dan 87,6%. Spiros V. Georgakopoulos, dkk[6] tentang klasifikasi teks dengan nilai akurasi 91,2%.

Berdasarkan penelitian sebelumnya, CNN memiliki kemungkinan memperoleh akurasi lebih baik dibandingkan algoritma lainnya jika diimplementasikan dalam ekstraksi informasi. Metode dengan menggunakan *full automatic training* tanpa harus membuat aturan perbaikan atau *ruled based*. Maka kali ini perlu dilakukan sebuah penelitian tentang ekstraksi informasi pada surat keputusan menggunakan algoritma *Convolutional Neural Network* (CNN).

## 1.2 Rumusan Masalah

Berdasarkan yang telah disebutkan pada latar belakang, maka rumusan masalahnya adalah berapa nilai akurasi, presisi, dan recall Algoritma CNN dalam implementasi ekstraksi informasi pada Surat Keputusan?

## 1.3 Maksud dan Tujuan Penelitian

Maksud dilakukannya penelitian ini adalah mengimplementasikan algoritma CNN dalam pembangunan sistem ekstraksi informasi pada surat keputusan.

Berdasarkan rumusan masalah yang telah dipaparkan di atas, maka tujuan dilakukannya penelitian ini adalah untuk mengetahui nilai akurasi, presisi, dan *recall* algoritma CNN dalam implementasi ekstraksi informasi pada surat keputusan.

## 1.4 Batasan Masalah

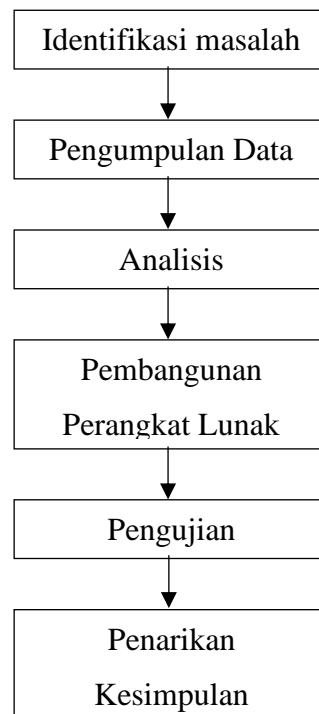
Batasan masalah pada penelitian yang akan dilakukan adalah:

1. Metode atau algoritma yang digunakan untuk mengekstraksi informasi adalah *Convolution Neural Network*.
2. Jenis dokumen teks yang digunakan untuk data training dan data testing pada penelitian ini berformat \*.pdf.
3. Data masukan yang digunakan pada penelitian ini adalah Surat Keputusan Pemerintah yang ditujukan untuk satu orang.

- Informasi yang diambil dari surat keputusan adalah Nomor surat, tentang, nama penerima, nip penerima, pangkat/golongan, unit kerja, tanggal ditetapkan, jabatan yang menetapkan, nama yang menetapkan, Null.

## 1.5 Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah metode penelitian deskriptif. Metode yang mempelajari masalah yang ada dengan tatacara kerja yang berlaku. Langkah-langkah penelitian dapat dilihat pada **Gambar 1.1 Sistematika Metode Penelitian Deskriptif**.



**Gambar 1.1 Sistematika Metode Penelitian Deskriptif**

### 1.5.1 Identifikasi Masalah

Berdasarkan latar belakang yang telah dipaparkan, permasalahan yang dapat diidentifikasi adalah bagaimana mengesktraksi informasi pada surat keputusan dengan menggunakan algoritma CNN.

### 1.5.2 Pengumpulan Data

Pengumpulan data dilakukan dengan cara mengambil data dari internet, jurnal, artikel ilmiah, ataupun dari sumber-sumber yang memiliki data-data yang akan dijadikan bahan dalam penelitian yang akan dilakukan.

Kerangka Pengumpulan data yang didapatkan dari objek penelitian yaitu :

1. Studi Pustaka

Studi yang dilakukan dalam mencari informasi yang terkait dengan penelitian melalui literatur jurnal, *paper*, *proceedings*, maupun buku.

2. Data Set

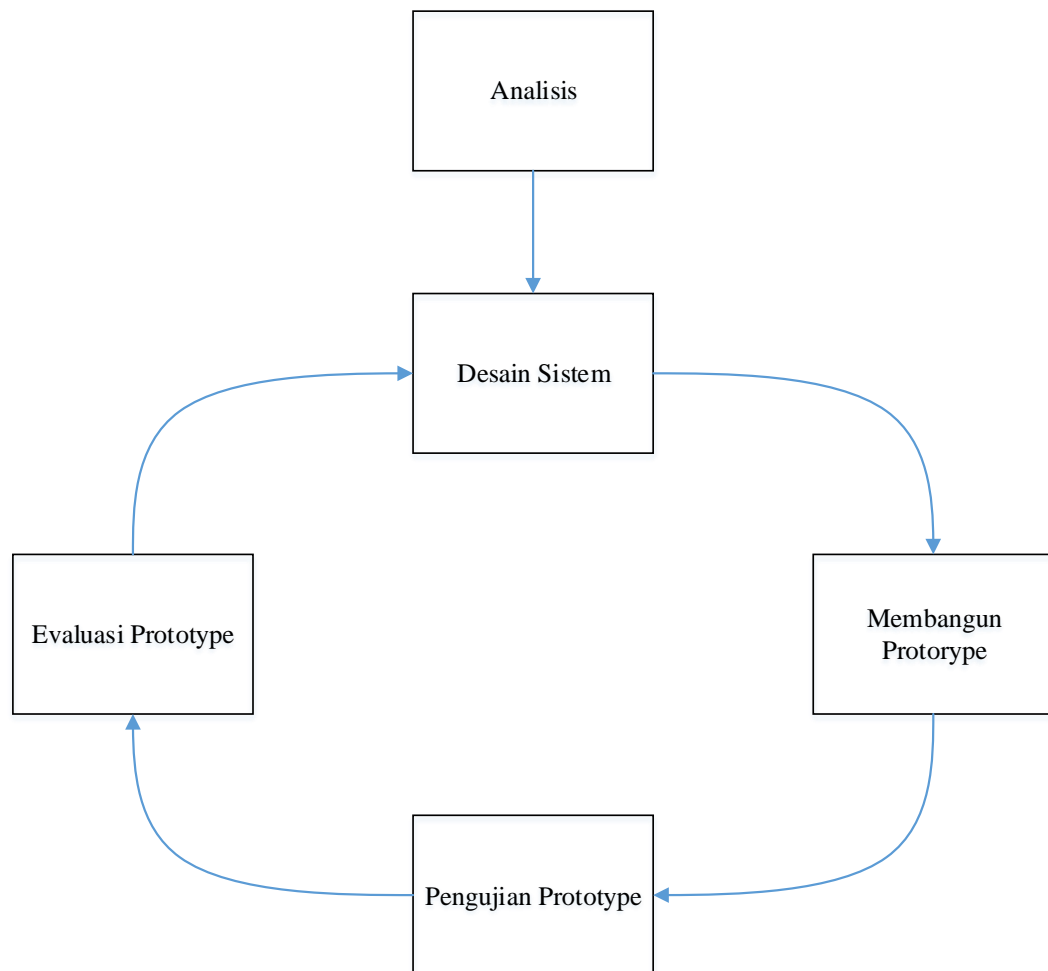
Pengumpulan data dengan mengumpulkan bahan dari internet, literatur, maupun sumber-sumber yang memiliki data yang akan menjadi bahan dalam penelitian.

### 1.5.3 Analisis

Tahap analisis yang dilakukan adalah analisis data masukan yang akan digunakan, analisis tahap *preprocessing*, sampai analisis ekstraksi dengan menggunakan metode CNN.

#### 1.5.4 Pembangunan Perangkat Lunak

Metode pembangunan perangkat lunak pada penelitian ini dengan metode *prototype*. Metode ini dapat dilihat pada **Gambar 1.2 Metode Pembangunan Perangkat Lunak**.



**Gambar 1.2 Metode Pembangunan Perangkat Lunak**

Pemodelan sistem yang digunakan menggunakan pemodelan *Prototype*. Berikut tahapan pada model *Prototype*.

1. Analisis Sistem

Mengumpulkan kebutuhan-kebutuhan untuk sistem yang dibuat kemudian dianalisis sesuai dengan kebutuhan penelitian seperti dokumen pendukung dan informasi tambahan yang dibutuhkan.

## 2. Desain

Menjabarkan informasi yang didapat dan telah dianalisis secara rinci seperti kebutuhan *software* dan *hardware* yang dimuat kedalam bentuk *blueprint* sebelum masuk ke tahap pembangunan *prototype* agar sistem dapat dibangun secara mendetail.

## 3. Pembangunan *Prototype*

Tahap penerjemahan desain ke dalam bahasa pemrograman yang berisikan pemecahan masalah pada penelitian.

## 4. Pengujian

Menguji apakah sistem yang dibangun dapat memecahkan masalah atau tidak, sesuai dengan apa yang telah dirancang sebelumnya.

## 5. Evaluasi *Prototype*

Pada tahap ini sistem akan mengalami perubahan-perubahan tergantung permintaan atau kebutuhan sistem.

### 1.5.5 Pengujian

Pengujian yang dilakukan pada penelitian ini menggunakan *confussion matrix* dengan menghitung *precision*, *recall*, dan akurasi.

### 1.5.6 Penarikan Kesimpulan

Penarikan kesimpulan dilakukan setelah pengujian dilakukan. Hasil dari pengujian dijadikan bahan untuk menarik kesimpulan, apakah CNN mempunyai hasil yang bagus atau tidak dalam implementasi ekstraksi informasi pada Surat Keputusan.

## 1.6 Sistematika Penulisan

Sistematika penulisan skripsi yang akan dibuat terbagi kedalam 5 Bab beserta pokok materinya. Sebagai gambaran umum, sistematika penyusunan skripsi yang akan ditulis adalah sebagai berikut :

## **BAB 1 PENDAHULUAN**

Pendahuluan menjelaskan mengenai Latar Belakang Masalah, Identifikasi Masalah, Maksud dan Tujuan Penelitian, Batasan Masalah, Metodologi Penelitian, dan Sistematika Penulisan.

## **BAB 2 TINJAUAN PUSTAKA**

Didalam bab ini membahas landasan teori yang menunjang dalam penelitian dalam menyelesaikan masalah.

## **BAB 3 ANALISIS DAN PERANCANGAN**

Pada bab ini membahas tentang analisis masalah dan kebutuhan mengenai metode *Convolutional Neural Network* dalam men ekstraksi dokumen. Penjelasan proses dalam ekstraksi informasi secara matematis. Pada bab ini juga dibahas tentang perancangan UML dan perancangan antarmuka.

## **BAB 4 IMPLEMENTASI DAN PENGUJIAN**

Pada bab ini membahas tentang implementasi dari hasil analisis dan perancangan yang telah dilakukan, penjelasan proses ekstraksi informasi secara sistematis, hasil penelitian, dan pengujian.

## **BAB 5 KESIMPULAN DAN SARAN**

Berisi kesimpulan dari hasil penelitian Ekstraksi Informasi pada Surat Keputusan Menggunakan Algoritma *Convolutional Neural Network* dan saran-saran untuk pengembangan penelitian yang akan datang.

