

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Minat penelitian terhadap kecerdasan buatan meningkat baru-baru ini. Menurut Ensiklopedia Britannica [1], kecerdasan buatan adalah satu dari cabang ilmu komputer yang dapat memproses informasi berdasarkan metode heuristic dengan sejumlah aturan. Kecerdasan buatan banyak diteliti dalam bidang kesehatan[2], pertanian[3], keuangan[4], hingga pada satelit telekomunikasi [5]. Dalam penelitian [6], dijelaskan bahwa kecerdasan buatan meliputi *Machine Learning*, dan *Deep Learning*. Hubungan kecerdasan buatan, *Machine Learning*, dan *Deep Learning* adalah dimana *Deep Learning* merupakan subbagian dari *Machine Learning*, dan *Machine Learning* subbagian dari kecerdasan buatan. Ketiga istilah tersebut memiliki identitas yang sama yaitu mesin yang diprogram untuk belajar dan menemukan solusi terbaik dalam usaha memecahkan suatu permasalahan.

*Machine Learning* merupakan bagian dari bidang kecerdasan buatan yang menggunakan sebuah komputer untuk membuat sebuah prediksi berdasarkan tren atau pola data sebelumnya [7]. *Machine Learning* adalah sebuah metode melatih komputer untuk menerapkan proses pembelajaran dari pengalaman di masa lampau dengan tujuan dapat memecahkan masalah yang diberikan. Hal ini memungkinkan *Machine Learning* untuk menganalisis dan memproses data yang besar untuk mendapatkan *insight* dan korelasi antardata [8]. Di sisi lain, dalam *Machine Learning* juga terdapat istilah *Deep Learning*. *Deep Learning* merupakan proses pembelajaran mesin yang terdiri atas banyak pengklasifikasi yang bekerja bersama, dan didasarkan pada regresi linier diikuti oleh beberapa fungsi aktivasi [9]. *Deep Learning* dapat diimplementasikan dalam banyak bidang, salah satunya pada *Natural Language Processing* (NLP). *Deep Learning* dapat diterapkan pada subbidang NLP, seperti analisis sentimen [10].

Dalam NLP, Analisis sentimen dapat diartikan sebagai proses menganalisis, mengolah dan memahami data teks dengan tujuan mendapatkan informasi

kecenderungan opini/pendapat dalam suatu permasalahan [11]. Penelitian yang membahas analisis sentimen, misalnya pada penelitian [12] yang bertujuan menyusun beberapa metode *preprocessing* untuk beberapa *dataset* pada komentar platform YouTube. Teknik pelabelan opini pada penelitian tersebut menggunakan *human annotation*. Sedangkan dalam penelitian [13], dijelaskan bahwa pelabelan manual menggunakan *human annotator* memiliki kelemahan terkait subjektivitas. Subjektivitas dapat diminimalisir dengan menambah jumlah annotator. Pada penelitian tersebut menggunakan 3 *annotator* yang berasal dari bidang ilmu bahasa, ilmu psikologi, dan teknik komputer. Hal tersebut tentu membutuhkan waktu yang lama dan biaya yang mahal.

Cara yang lain adalah dengan menggunakan bantuan komputer (*machine annotator*). Dalam penelitian [14] yang menggunakan metode *sentiment lexicons* sebagai *machine annotator*, dijelaskan bahwa *sentiment lexicons* memiliki permasalahan dimana dalam media sosial (misal: Twitter) manusia lebih memilih menggunakan kosakata informal dan emoji untuk menyampaikan emosi, daripada menggunakan kosakata formal seperti pada kamus *Lexicon*. Selain itu, hubungan antara kata-kata dan emosi bervariasi dari satu domain ke domain lain, ini lebih dikenal dengan disambiguasi kontekstual.

Berdasarkan permasalahan yang telah dijabarkan, diperlukan sebuah solusi dimana anotasi dapat dilakukan otomatis sehingga dapat meminimalisir terjadinya disambiguasi kontekstual, sekaligus dapat mengurangi durasi dalam proses anotasi. Maka, penelitian ini bertujuan mengusulkan pemodelan anotasi (pelabelan) “Pelabelan Sentimen Berbasis Semi-Supervised Learning menggunakan *Deep Learning*”. SSL merupakan gabungan antara anotasi menggunakan *human annotator* dan *machine annotator*. Dalam penelitian ini menggunakan *machine annotator* berupa algoritma *Deep Learning* yaitu algoritma Long Short-Term Memory (LSTM) dan algoritma Gated Recurrent Unit (GRU). LSTM merupakan algoritma turunan dari Recurrent Neural Network (RNN) yang banyak digunakan dalam penelitian penambangan data teks. Alasan digunakannya algoritma LSTM adalah kemampuan algoritma LSTM dalam memproses data sekuensial yang panjang sehingga mampu mengatasi

permasalahan *vanishing gradient* [15], dan termasuk algoritma dengan tingkat akurasi klasifikasi cenderung lebih baik [16], [17]. Algoritma GRU merupakan algoritma *Deep Learning* yang memiliki struktur model mirip dengan Algoritma LSTM dengan versi yang lebih sederhana sehingga memiliki waktu komputasi yang lebih singkat. Ini ditunjukkan oleh penelitian [18], [19] dengan membandingkan algoritma GRU dan LSTM dengan waktu komputasi GRU yang lebih singkat daripada LSTM pada seluruh skenario.

## 1.2 Rumusan Masalah

Pelabelan sentimen analisis secara manual oleh manusia (*human annotator*) memiliki kelemahan terkait subjektivitas, dan waktu yang lama untuk data yang banyak. Sedangkan pelabelan menggunakan bantuan komputer (*machine annotator*) sering terdapat disambiguasi kontekstual. Maka diperlukan komputasi otomatis yang mempelajari pola pelabelan oleh manusia.

## 1.3 Pertanyaan Penelitian

Berdasarkan rumusan masalah diatas, maka pertanyaan penelitian antara lain:

1. bagaimana implementasi metode pelabelan sentimen berbasis *Semi-Supervised Learning* (SSL) menggunakan *Deep Learning*?
2. berapa tingkat akurasi SSL metode pelabelan sentimen berbasis *Semi-Supervised Learning* (SSL) menggunakan *Deep Learning*?

## 1.4 Batasan Masalah

Batasan masalah pada penelitian ini antara lain:

1. penelitian ini terbatas pada *dataset* komentar *review e-wallet* DANA, dan OVO, untuk data tanpa label menggunakan *review* aplikasi GoPay dari Google PlayStore.
2. penelitian ini terbatas pada rating yang digunakan pada *dataset* yaitu rating 1, 2, 4, dan 5.
3. penelitian ini terbatas pada model *binnary classification*.

### **1.5 Tujuan Penelitian**

1. Mengetahui implementasi metode pelabelan sentimen berbasis *Semi-Supervised Learning* (SSL) menggunakan *Deep Learning*.
2. Mengetahui tingkat akurasi SSL metode pelabelan sentimen berbasis *Semi-Supervised Learning* (SSL) menggunakan *Deep Learning*.

### **1.6 Manfaat Penelitian**

1. Manfaat teoritis, informasi yang dihasilkan oleh model anotasi dapat digunakan sebagai ilmu pengetahuan dalam melakukan pengembangan model *Semi-Supervised Learning* (SSL).
2. Manfaat praktis, dapat dipergunakan sebagai bahan referensi, pengembangan maupun pembandingan dalam penelitian selanjutnya.
3. Manfaat bagi peneliti, dapat mempelajari ilmu pengetahuan dan informasi baru serta pengalaman dan keterampilan mengolah data teks atau *Natural Language Processing*.