

# BAB 1 PENDAHULUAN

## 1.1 Latar Belakang

Di era globalisasi saat ini, pengguna Internet dapat dengan mudah mencari informasi mengenai suatu hal. Selain itu, pengguna juga dapat menuliskan opini mereka mengenai sebuah produk atau jasa. Salah satu informasi yang terdapat di Internet adalah informasi mengenai buku. Pengguna Internet dapat dengan mudah menuliskan opini atau ulasan mereka mengenai sebuah buku. Dengan membaca *review* atau ulasan berdasarkan pengalaman pembaca lain, maka kita akan mengetahui kualitas dari buku tersebut. Namun, begitu banyaknya opini akan mempersulit pengguna lain untuk memperoleh informasi dari opini tersebut.

Solusi dari permasalahan tersebut adalah dengan menggunakan klasifikasi sentimen. Klasifikasi sentimen merupakan salah satu bagian dari klasifikasi teks, dimana sistem akan melakukan pengklasifikasian sentimen atau opini menjadi sentimen positif atau negatif [1]. Dengan melakukan pengklasifikasian teks pada opini ulasan buku, akan membantu pembaca memperoleh informasi kualitas buku secara cepat.

Dalam klasifikasi sentimen, terdapat masalah ketidakpastian atau *uncertainty reasoning* pada teks. Sebagai contoh apabila terdapat dua kalimat yang memiliki fitur atau kata yang sama, sistem akan mengalami kesulitan dalam proses pengklasifikasian [2]. Masalah tersebut dapat diselesaikan dengan menggunakan salah satu metode pada pengklasifikasian teks yaitu *Naïve Bayes* [2]. *Naïve Bayes* merupakan metode yang menerapkan *Bayesian theorem*, dimana *Naïve Bayes* mengasumsikan secara *naive* (sederhana) bahwa setiap fitur bersifat independensi, maksudnya adalah setiap fitur dalam data yang sama tidak saling berkaitan [3]. Selain itu, *Naïve Bayes* merupakan metode yang melihat pendapat atau opini tersebut, cenderung beropini positif atau negatif berdasarkan nilai probabilitasnya [4]. *Naïve Bayes* juga memiliki performa yang tinggi dengan kalkulasi yang

sederhana [2]. Berdasarkan hal tersebut, penulis menggunakan *Naïve Bayes* pada penelitian klasifikasi sentimen ulasan buku ini.

Permasalahan lain dalam klasifikasi sentimen adalah jika terdapat terlalu banyak jumlah fitur pada suatu *dataset*, akan menyebabkan turunnya akurasi dari klasifikasi [5]. Oleh karena itu, dibutuhkan *feature selection* yang digunakan untuk memilih fitur yang signifikan, dan membuang fitur yang tidak relevan, sehingga dimensi data akan berkurang dan akurasi akan meningkat [6]. *Information Gain* merupakan salah satu teknik pada *feature selection*, dimana suatu kata diukur dengan menghitung jumlah informasi kata tersebut ada atau tidak pada suatu dokumen [7]. Pada penelitian ini, *Information Gain* digunakan sebagai *feature selection* untuk memilih fitur-fitur yang signifikan dan menghilangkan fitur-fitur yang tidak penting sehingga proses klasifikasi sentimen ulasan buku lebih akurat.

## 1.2 Perumusan Masalah

Adapun masalah yang akan diteliti adalah sebagai berikut:

1. Bagaimana implementasi dari kombinasi metode *Information Gain* pada *feature selection* dan *Naïve Bayes* pada klasifikasi sentimen ulasan buku?
2. Berapa performa yang dihasilkan dari penelitian klasifikasi pada ulasan buku?

Adapun batasan masalah yang digunakan pada penelitian ini adalah

1. Data diambil dari *Goodreads* [8]. Data tersebut berisi ulasan atau *review* buku. *Dataset* tersebut telah diberi label berupa 2 *class* yaitu positif dan negatif.
2. Data ulasan buku menggunakan bahasa Inggris.

### 1.3 Tujuan

Tujuan dari penelitian Tugas Akhir ini adalah sebagai berikut:

1. Mengimplementasikan kombinasi *Information Gain* untuk *feature selection* dan *Naïve Bayes* untuk klasifikasi sentimen ulasan buku.
2. Mengetahui performa dari penerapan klasifikasi sentimen menggunakan *Information Gain* dan *Naïve Bayes*.

### 1.4 Metodologi Penyelesaian Masalah

Metodologi pada penelitian ini adalah sebagai berikut

1. Studi Literatur  
Penelitian ini dimulai dengan proses pengumpulan referensi mengenai analisis sentimen terhadap ulasan produk atau jasa dan metode klasifikasi *Naïve Bayes*.
2. Pengumpulan *Dataset*  
Pada tahap ini, penulis melakukan pengumpulan *dataset* dari situs buku *Goodreads*. Data yang dikumpulkan sebanyak 1000 *record* data ulasan buku, yang terdiri dari dua *class* yaitu positif dan negatif.
3. Analisis & Perancangan Sistem  
Pada tahap ini, penulis melakukan klasifikasi sentimen pada ulasan buku berbahasa Inggris. Sistem dibangun menggunakan *Information Gain* sebagai *feature selection* dan *Naïve Bayes* untuk proses klasifikasi.
4. Implementasi Sistem  
Pada tahap ini, penulis mengimplementasikan sistem yang dibangun dengan menggunakan *Information Gain* dan *Naïve Bayes*.
5. Pengujian, analisis, dan evaluasi  
Hasil pengujian akan dianalisis serta dievaluasi untuk menyimpulkan hasil penelitian. Pada tahap ini, evaluasi yang digunakan adalah *confusion matrix* dan *5-fold-cross validation*.
6. Penyusunan laporan tugas akhir

Tahap ini merupakan tahap dimana, penulis melakukan pembuatan laporan akhir sebagai bentuk dari dokumentasi terkait penelitian ini.

## **1.5 Sistematika Penulisan**

Sistematika penulisan dalam Buku Tugas Akhir ini terdiri dari lima bab, dengan rincian sebagai berikut:

### **BAB 1 PENDAHULUAN**

Bab ini menjelaskan tentang latar belakang masalah, perumusan masalah, batasan masalah, tujuan, metodologi dan sistematika penulisan.

### **BAB 2 KAJIAN PUSTAKA**

Bab ini menjelaskan tentang dasar teori yang mendukung dalam penelitian ini. Teori tersebut diambil dari beberapa literatur terkait.

### **BAB 3 PERANCANGAN SISTEM**

Bab ini menjelaskan tentang gambaran umum sistem, tahapan tiap proses penelitian, dan spesifikasi lingkup operasional yang digunakan pada penelitian.

### **BAB 4 PENGUJIAN DAN ANALISIS**

Bab ini berisi tentang hasil pengujian dan analisis dari hasil penelitian yang telah dilakukan.

### **BAB 5 KESIMPULAN DAN SARAN**

Bab ini berisi tentang kesimpulan dan saran yang berkaitan dengan hasil penelitian.