BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Bahasa dalam sebuah tulisan merupakan alat komunikasi untuk menerjemahkan sebuah pikiran yang ingin diungkapkan oleh seorang penulis. Dalam sebuah tulisan seorang penulis tidak hanya menyampaikan sebuah keterangan dari suatu informasi, tetapi juga berisi informasi tentang perilaku manusia termasuk emosi [1]. Emosi yang tersembunyi dibalik tulisan sangat sulit untuk ditafsirkan oleh pembaca dikarenakan pengaruh oleh sudut pandang yang berbeda antara penulis dan pembaca [2]. Maka dari itu untuk memudahkan pembaca mengetahui emosi dari sebuah tulisan bisa menggunakan cara klasifikasi.

Banyak metode yang bisa digunakan dalam klasifikasi teks diantaranya yaitu metode *naive bayes*, *k-nearest neighbor* (KNN), *support vector machine* (SVM) *decision tree, neural network* dan sebagainya. Masing – masing metode mempunyai kelebihan dan kekurangan masing – masing dan tingkat akurasi yang berbeda-beda. Seperti pada penelitian yang berjudul "Review on Comporation Between Text Classification Algorithms" dari hasil perbandingan algoritma *naive bayes*, *k-nearest neighbor* (KNN), dan *support vector machine* (SVM), diketahui bahwa algoritma *K-nearest neighbor* (KNN) memiliki kinerja yang bagus dengan nilai akurasi sampai 86% [3].

Pada penelitian yang sudah dilakukan sebelumnya dengan judul "Klasifikasi Emosi Untuk Teks Berbahasa Indonesia Dengan Menggunakan K-Nearest Neighbor" dijelaskan tahapan untuk mengenali emosi pada lirik lagu dilakukan beberapa proses. Adapun tahapan yang dilakukan sebelum proses klasifikasi dengan *k-nearest neighbor* (KNN) adalah tahapan *preprocessing* dan tahapan pembobotan dengan menggunakan metode TF.IDF, hasil akurasi yang didapatkan adalah sekitar 50% – 60% [4]. Hasil tersebut dipengaruhi oleh metode pembobotan yang digunakan, seperti dalam penelitian "Text Categorization Based on Weighted Inverse Document Frequency" menjelaskan pada pembobotan term dengan metode

TF.IDF mempunyai kelemahan yaitu semua *text* yang berisi *term* tertentu diperlakukan sama, sehingga IDF tidak memperhitungkan jumlah kemunculan suatu *term* pada suatu dokumen. Sehingga untuk metode pembobotan term yang akan digunakan dalam penelitian ini adalah metode WIDF (*Weighted Inverse Document Frequency*), metode ini memiliki kinerja yang lebih baik dibanding dengan TF.IDF dengan memberikan peningkatan akurasi pada IDF sebesar 7.4% Maksimum [5].

Berdasarkan gambaran yang telah dijelaskan, dalam penelitian ini digunakan metode *K-nearest neighbor* (KNN) dengan metode pembobotan *term* WIDF untuk mengklasifikasikan emosi pada teks bahasa Indonesia dan mengetahui besar akurasi yang didapatkan.

1.2 Perumusan Masalah

Berdasarkan penjelasan dari latar belakang masalah, maka rumusan masalah yang dapat teridentifikasi adalah bagaimana *K-nearest neighbor* dengan pembobotan WIDF dalam mengklasifikasikan emosi pada sebuah teks bahasa Indonesia dan mengetahui seberapa besar akurasi yang didapatkan.

1.3 Maksud dan Tujuan

Berdasarkan permasalahan yang akan diteliti, maka maksud penelitian ini adalah mengklasifikasikan emosi pada teks bahasa Indonesia menggunakan metode *K-nearest neighbor* dengan pembobotan WIDF

Tujuan yang ingin dicapai dalam simulasi pengenalan emosi pada teks bahasa Indonesia adalah :

1. Untuk mengklasifikasikan emosi pada teks bahasa Indonesia
2. Untuk mengetahui besar akurasi yang didapatkan dari pengklasifikasian emosi pada teks bahasa Indonesia menggunakan metode *K-nearest neighbor* dengan pembobotan WIDF.

1.4 Batasan Masalah

Ada pun batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Data latih digunakan adalah dari ISEAR *(International Survey On Emotion Antecedents And Reaction)*
2. Data uji yang digunakan adalah lirik lagu dengan berformat (.txt)

3. Kategori emosi senang, takut, marah, sedih dan bersalah yang disesuaikan dengan penelitian sebelumnya.

1.5 Metodologi Penelitian

Metodologi penelitian merupakan sekumpulan peraturan, kegiatan dan prosedur yang digunakan oleh peneliti untuk memecahkan suatu masalah agar lebih efisien. Metode penelitian yang digunakan adalah metode deskriptif. Metode deskriptif bertujuan untuk mendapatkan gambaran yang jelas tentang hal-hal yang diperlukan, melalui tahapan sebagai berikut:

1.5.1 Metode Pengumpulan Data

Dalam tahapan ini pengumpulan data yang dilakukan adalah :

a. Studi Literatur

Pada tahap ini akan dilakukan pengumpulan informasi yang berhubungan dengan Tugas Akhir ini khususnya pada metode yang diterapkan *K-Nearset Neighbor* (KNN) melalui sumber-sumber seperti jurnal penelitian, buku-buku teori, dan sumber-sumber lain.

b. Observasi

Observasi merupakan metode pengumpulan data dengan cara mengamati objek yang berkaitan dengan penelitian secara langsung terhadap permasalahan yang diambil. Di mana observasi yang dilakukan adalah dengan menggunakan data dari ISEAR (International Survey On Emotion Antecedents And Reaction) untuk data latihnya dan data uji mengggunakan lirik lagu yang akan diteliti.

1.5.2 Metode Pembangunan Perangkat Lunak

Metode pembangunan perangkat lunak merupakan langkah – langkah yang akan dilaksanakan dalam membangun simulasi untuk mengklasifikasikan emosi pada teks bahasa Indonesia dengan pembobotan WIDF. Berikut tahapan metode yang akan dilakukan dalam penelitian ini :

1. Analisis Metode

   Dalam tahapan ini dilakukan untuk mempelajari konsep dan analisi mengenai klasifikasi teks, metode *K-Nearest Neighbor*, dan pembobotan WIDF yang akan diimplementasikan dalam sistem

2. Perancangan Desain

   Pada tahap ini dilakukan perancangan desain dari simulasi klasifikasi emosi pada teks bahasa indonesia yang akan dibuat dengan analisis yang telah dilakukan pada tahap sebelumnya.

3. Pembangunan simulasi

   Pada tahap ini dilakukan proses implementasi metode K-Nearest Neighbor dengan pembobotan WIDF ke dalam simulasi yang akan dibangun sesuai dengan desain yang telah dibuat pada tahap sebelumnya.

4. Pengujian simulasi

   Pada tahap ini simulasi yang telah dibangun akan diuji oleh penguji, penguji menilai apakah simulasi yang dibangun sudah sesuai dengan tujuan dari penelitian.

Adapun gambaran dari tahapan pembangunan perangkat lunak yang akan dilakukan dalam penelitian ini dapat dilihat pada gambar 1.1



Gambar 1.1 Tahapan Pembangunan Perangkat Lunak

1.6 Sistematika Penulisan

Penulisan skripsi ini terbagi menjadi beberapa bab yang masing – masing bab membahas tentang :

BAB 1 PENDAHULUAN

Bab ini menjelaskan tentang latar belakang masalah, perumusan masalah, maksud dan tujuan, menentukan batasan masalah, serta menjelaskan mengenai metode penelitian dan sistematika penulisan.

BAB 2 LANDASAN TEORI

Di dalam bab ini membahas tentang konsep dasar dan teori-teori yang berkaitan dengan topik penelitian yang dilakukan dan hal-hal yang berguna dalam proses analisis permasalahan serta tinjauan terhadap penelitian.

BAB 3 ANALISIS DAN PERANCANGAN SISTEM

Bab ini menjelaskan analisis dan perancangan sistem. Analisis yang dilakukan berupa analisis masalah, analisis penyelesaian masalah, analisis data masukan, analisis kebutuhan non fungsional dan analisis kebutuhan fungsional. Perancangan yang dilakukan berupa perancangan data, perancangan antarmuka, dan perancangan prosedural.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Bab ini menjelaskan mengenai implementasi dari hasil tahapan analisis dan perancangan sistem yang dibangun. Serta berisi pengujian perangkat lunak seperti pengujian *black box*, pengujian *precision, recall*.

BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan yang didapatkan selama penelitian yang dilakukan, selain itu juga berisi saran untuk perbaikan dan menindaklanjuti hasil penelitian.