

BAB I PENDAHULUAN

1.1 Latar Belakang

Evaluasi merupakan salah satu kegiatan penting dalam proses pembelajaran, karena dengan melakukan evaluasi dapat diketahui hasil belajar dan keefektifan proses pembelajaran tersebut. Hasil evaluasi dapat digunakan untuk menentukan langkah-langkah apa yang perlu dilakukan untuk memperbaiki pelaksanaan pembelajaran selanjutnya [NUG 13]. Bentuk evaluasi yang biasa digunakan adalah tes atau ujian.

Kenyataannya, dalam menghadapi tes tersebut banyak siswa yang merasa kesulitan. hal ini dapat disebabkan oleh proses pembelajaran yang kurang optimal. Di kelas, seorang guru mengajar dengan pola yang sama terhadap semua siswa, sehingga kurang efektif karena kemampuan, karakteristik, dan pengetahuan setiap siswa tidak sama.

Dalam penelitian yang dilakukan oleh Dwi Ratnawati (2011), memberikan solusi berupa penerapan Sistem Tutor Cerdas (STC). STC adalah suatu *software* yang dapat menyediakan referensi materi pembelajaran yang sesuai dengan pengguna, sehingga proses pembelajaran berlangsung lebih efektif dan efisien. Salah satu fitur yang dimiliki STC tersebut adalah mampu mengklasifikasikan soal yang berupa kumpulan teks ke dalam kelas-kelas yang sudah ditentukan. Soal-soal yang sudah diklasifikasikan dapat dikembangkan berdasarkan level kesulitan dengan mendefinisikan aturan-aturan yang ada, sehingga dapat membantu seorang pengajar dalam mengevaluasi siswanya, sekaligus dapat memberikan materi-materi pelajaran tertentu secara interaktif. [RAT 11].

Metode *K-Nearest Neighbor* (KNN) adalah salah satu metode yang dapat digunakan untuk menyelesaikan masalah klasifikasi teks. Teknik ini sangat sederhana dan mudah diimplementasikan. Mirip dengan teknik *clustering*, mengelompokkan suatu data baru berdasarkan jarak baru itu ke beberapa data atau tetangga (*neighbor*) terdekat. KNN adalah suatu metode pengklasifikasian berdasarkan mayoritas dari kategori. Metode ini bertujuan untuk mengklasifikasi objek baru berdasarkan atribut dan *training sample*. Metode ini memiliki kelebihan seperti efektifitas yang baik dan apabila *training data* berjumlah besar

akan semakin efektif, juga mempunyai beberapa kekurangan diantaranya perlu ditentukan nilai k yang paling optimal yang menyatakan jumlah tetangga terdekat dan biaya komputasi yang cukup tinggi.

Untuk memperbaiki kekurangan yang ada pada metode KNN dapat digunakan metode *improved* KNN. Penelitian yang dilakukan oleh Li Baoli, Yu Shiwen, dan Lu Qin, mengklasifikasikan data berupa dokumen berbahasa China dengan metode *improved* KNN kedalam 12 kategori diantaranya *Humanities and Art*, *News and Media*, *Entertainment*, dan *Education*. Hasil penelitian menunjukkan bahwa metode *Improved* KNN rata-rata memiliki performansi lebih tinggi 1,4% dibandingkan KNN, selain itu didapatkan kestabilan proses kategorisasi walaupun menggunakan nilai k -values yang berbeda-beda [BAO 03].

Penelitian lain yang membahas tentang *improved* KNN adalah penelitian yang dilakukan oleh Shengyi Jiang, Guansong Pang, Meiling Wu, dan Limin Kuang [JIA 12].

Data yang digunakan berasal dari dua sumber, yaitu dokumen *Reuters-21578* yang berisi 21.578 dokumen dengan 135 kategori dan *Fudan Univ. Text categorization corpus* dengan 9804 dokumen training dan 9833 dokumen testing dengan 20 kategori.

Hasil penelitian untuk dokumen *Reuters-21578* menunjukkan bahwa *improved* K-NN memiliki performansi lebih baik sekitar 2,16% daripada K-NN sedangkan untuk dokumen *Fudan Univ. Text categorization corpus* 9804 didapatkan hasil lebih baik, yaitu metode *improved* K-NN dengan performansi lebih tinggi 4,13% lebih baik dari K-NN [JIA 12].

Metode *improved* K-NN telah diimplementasikan pada dokumen dengan bahasa China, tetapi seharusnya metode tersebut juga bekerja dengan baik untuk bahasa Indonesia. Berdasarkan latar belakang tersebut, penulis tertarik untuk membuat suatu sistem yang mampu mengklasifikasikan soal berdasarkan topiknya dengan menggunakan metode *improved* K-NN.

1.2. Rumusan Masalah

Berdasar latar belakang di atas, maka rumusan masalah pada skripsi ini adalah sebagai berikut :

1. Bagaimana menerapkan metode *Improved* KNN untuk mengklasifikasi soal-soal biologi SMA berdasarkan topik.
2. Bagaimana performa metode *Improved* KNN untuk klasifikasi soal-soal biologi SMA.

1.3. Batasan Masalah

Batasan masalah yang digunakan agar permasalahan yang dibahas lebih fokus adalah sebagai berikut :

1. Soal-soal yang akan diklasifikasikan adalah soal mata pelajaran Biologi SMA Kelas 10.
2. Soal-Soal akan diklasifikasikan berdasarkan topik bahasan.
3. Topik yang digunakan adalah tumbuhan, hewan, ekosistem, dan Protista. Data latih yang digunakan diambil dari beberapa buku pelajaran biologi yaitu :
 - a. Biologi SMA dan MA untuk Kelas X esis [ARY 07].
 - b. Biologi SMA dan MA untuk Kelas X Erlangga [PRA 12].
 - c. Biologi untuk Kelas X SMA dan MA Depdiknas [SUB 09].
 - d. Praktis Belajar Biologi untuk Kelas X SMA/MA Depdiknas [FER 09].

1.4. Tujuan

Tujuan yang ingin dicapai dalam penelitian ini adalah :

1. Menerapkan text mining dengan metode klasifikasi *improved* K-Nearest Neighbour untuk mengklasifikasikan soal-soal biologi sekolah menengah atas sesuai dengan topiknya.
2. Mengetahui performa klasifikasi dengan metode klasifikasi *improved* K-Nearest Neighbour untuk mengklasifikasikan soal-soal biologi.

1.5. Manfaat

Dengan penelitian ini diharapkan dapat membantu guru dalam penyusunan soal-soal tes yang memenuhi standar. Dengan adanya sistem ini diharapkan siswa dapat terbantu dalam belajarnya. Selain itu, penelitian ini dapat menjadi referensi dan pembelajaran bagi pihak lain yang melakukan penelitian serupa.

1.6. Sistematika Pembahasan

1. BAB I PENDAHULUAN

Bab ini berisi latar belakang penulisan, permasalahan yang dihadapi, batasan masalah, tujuan dan manfaat serta sistematika penulisan skripsi.

2. BAB II KAJIAN PUSTAKA

Bab ini berisi tentang dasar - dasar teori yang akan dipergunakan sebagai landasan utama pembuatan aplikasi.

3. BAB III METODOLOGI DAN PERANCANGAN

Bab ini berisi tentang metode atau langkah - langkah yang digunakan dalam penelitian skripsi dan perancangan sistem yang akan dibangun .

4. BAB IV IMPLEMENTASI

Bab ini berisi tentang implementasi sistem.

5. BAB V ANALISIS DAN PENGUJIAN

Bab ini berisi tentang pengujian dan analisis sistem yang dibangun.

6. BAB VI PENUTUP

Bab ini berisi tentang kesimpulan dan saran.