

BAB I PENDAHULUAN

I.1 Latar Belakang

Perkembangan teknologi informasi di dunia saat ini semakin tak terbendung. Teknologi informasi sudah memasuki berbagai bidang, salah satunya adalah bidang Pendidikan. Salah satu konsep dari teknologi informasi yang sangat bermanfaat di bidang Pendidikan adalah *E-Learning*. *E-Learning* merupakan suatu bentuk model pembelajaran dan didukung dengan menggunakan teknologi informasi dan komunikasi (Hanum, 2013). Dari segi siswa, dengan adanya *E-learning* siswa dapat langsung mengakses mata pelajaran dan mencari informasi terkait pelajaran secara online tidak terpaku pada sebuah ruang kelas. Dari segi guru, *E-learning* dapat membantu guru dalam mengelola peserta didik, memberi penugasan, diskusi, bahkan memberikan penilaian tanpa harus bertatap muka secara langsung (Abror & Jati, 2016). Di dalam *E-Learning* terdapat berbagai materi dan kumpulan soal-soal yang dapat diakses kapan saja dan dimana saja.

Pada tanggal 02 Maret 2020 merupakan awal munculnya virus COVID-19 di Indonesia. Pada saat itu penyebaran virus COVID-19 semakin tinggi di Indonesia. Banyak sekali kegiatan masyarakat yang semula luring dialihkan menjadi daring, untuk mengurangi penyebaran virus COVID-19. Salah satu sektor yang berubah dari luring menjadi daring di saat pandemi COVID-19 adalah sektor pendidikan. Pada saat pandemi COVID-19 terjadi di Indonesia, itu menjadi momentum maraknya penggunaan sistem *E-Learning*. Permasalahan yang datang adalah siswa-siswa baru menggunakan *E-Learning* sehingga belum terbiasa (Ichsan et al., 2020). Ada beberapa faktor yang mendukung kesiapan penerapan *E-Learning* yaitu (1) kesiapan peserta didik (2) kesiapan guru (3) infrastruktur (4) dukungan *management* (5) budaya sekolah (6) kecenderungan pembelajaran tatap muka (Jamal, 2020). Butuh banyak sekali dukungan dari pihak terkait agar *E-Learning* bisa diterapkan secara maksimal sebagai model pembelajaran baru.

Kegiatan belajar mengajar di sekolah merupakan kegiatan yang paling mendasar (Yulistiawati, 2019). Pendidikan merupakan hal yang penting bagi seluruh manusia. Ada beberapa daerah yang sudah menerapkan peraturan wajib belajar 12 tahun yaitu dari Sekolah Dasar (SD) sampai Sekolah Menengah Atas (SMA) atau

Sekolah Menengah Kejuruan (SMK). Siswa kelas 10 SMA merupakan siswa yang masih dalam tahap transisi dari Sekolah Menengah Pertama (SMP) dan masih berfokus pada penjurusan kelas. Siswa kelas 12 SMA merupakan siswa yang sudah fokus pada Ujian Nasional (UN) dan sudah fokus pada mata pelajaran yang dipilihnya saja seperti Fisika, Ekonomi, Biologi, dan yang lainnya. Sedangkan kelas 11 merupakan kelas pertengahan yang tidak lagi berfokus pada penjurusan kelas dan UN (Winarni et al., 2015). Pada jenjang SMA dibagi menjadi dua jurusan yaitu Ilmu Pengetahuan Alam (IPA) dan Ilmu Pengetahuan Sosial (IPS). Pada jurusan IPA terdapat mata pelajaran wajib yang harus dipelajari yaitu Matematika, Fisika, Biologi, dan Kimia. Biologi adalah salah satu mata pelajaran yang memiliki berbagai macam istilah latin dan materi yang kompleks yang terkadang membuat siswa merasa sulit memahaminya, biologi juga memiliki cakupan topik materi yang luas seperti manusia, hewan, dan tumbuhan (Yulistiawati, 2019). Buku Siap Pintar Belajar Mandiri (SPBM) merupakan suatu buku untuk pelajar atau guru SMA yang berisi materi dan soal-soal yang lengkap dan mudah dipahami yang memuat mata pelajaran penting seperti Biologi dan buku ini juga terdapat varian untuk setiap kelasnya.

Klasifikasi merupakan suatu proses untuk mengelompokkan objek yang memiliki karakteristik atau ciri yang sama kedalam beberapa kelas (Widiastuti et al., 2017). Klasifikasi topik soal merupakan cara mengelompokkan soal-soal berdasarkan kategori topik atau Bab, jadi soal akan dikelompokkan berdasarkan karakteristik Bab yang sama. Dalam melakukan klasifikasi topik soal diperlukan data yang banyak dan beranekaragam, sehingga dibutuhkannya proses *data mining*. Proses klasifikasi topik soal dapat dilakukan secara otomatis menggunakan aplikasi data mining. Dengan dibuatnya suatu sistem yang dapat melakukan klasifikasi soal berdasarkan topik soal atau Bab secara otomatis pada *platform E-learning*, yang nantinya dapat memudahkan para siswa dan guru dalam melakukan kegiatan belajar mengajar yaitu dalam menentukan topik soal dari suatu soal.

Ada berbagai macam metode klasifikasi, diantaranya adalah algoritma Naive Bayes dan algoritma C4.5. Algoritma Naive Bayes dan algoritma C4.5 merupakan metode klasifikasi pada *text mining*. Pada penelitian yang dilakukan dalam

(Kamalia et al., 2019) yang membandingkan algoritma Naive Bayes, Support Vector Machine, C4.5, dan K-Nearest Neighbour. Pada penelitian mereka, data yang digunakan adalah ulasan pada aplikasi bibit yang ada di *play store*. Data yang digunakan terdiri dari 464 *dataset* yang terbagi menjadi 310 data sentimen positif dan 154 sentimen negatif. Hasil dari penelitian mereka adalah algoritma Naive Bayes memiliki nilai akurasi tertinggi sebesar 84.91%, C4.5 sebesar 76.94%, Support Vector Machine sebesar 71.77%, dan K-Nearest Neighbour sebesar 66,81. Adapun penelitian yang dilakukan dalam (Sari, 2017) yang membandingkan algoritma Support Vector Machine, Naive Bayes, dan C4.5. Data yang digunakan adalah *Short Message Service* (SMS) spam dan non spam. dataset terdiri dari 100 SMS spam dan 100 SMS non spam. Hasil dari penelitian tersebut adalah didapat nilai akurasi tertinggi oleh C4.5 sebesar 95.5%, Naive Bayes sebesar 95%, sedangkan Support Vector Machine sebesar 76%.

Algoritma Naive Bayes dan C4.5 digunakan karena dianggap cukup mudah dipahami dan merupakan algoritma yang sangat populer dalam melakukan klasifikasi. Naive Bayes merupakan algoritma klasifikasi dengan rumus yang sederhana dan mudah untuk diaplikasikan, serta mampu menghasilkan klasifikasi yang akurat (Maryamah et al., 2016). Sedangkan algoritma C4.5 dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima, efisien dalam menangani atribut bertipe diskret dan dapat menangani atribut bertipe diskret dan numerik (Kamagi & Hansun, 2014). Sehingga pada tugas akhir ini peneliti bermaksud melakukan perbandingan hasil akurasi antara algoritma Naive Bayes dan algoritma C4.5 dalam melakukan klasifikasi topik soal pada mata pelajaran biologi SMA kelas 11 dengan menggunakan *dataset* soal biologi yang terdapat pada buku SPBM kelas 11 SMA.

I.2 Perumusan Masalah

Rumusan masalah yang mendasari penelitian ini adalah:

- a. Bagaimana hasil akurasi dari klasifikasi soal biologi SMA kelas 11 berdasarkan kategori topik dengan implementasi algoritma Naive Bayes dan C4.5?

- b. Bagaimana hasil evaluasi performansi dari klasifikasi soal biologi SMA kelas 11 berdasarkan kategori topik dengan implementasi algoritma Naive Bayes dan C4.5?
- c. Bagaimana perbandingan tingkat akurasi dan evaluasi performansi dari algoritma Naive Bayes dan C4.5 dalam mengklasifikasikan soal biologi SMA kelas 11 berdasarkan kategori topik?

I.3 Tujuan Penelitian

Penelitian ini bertujuan untuk:

- a. Mengetahui cara mendapatkan hasil nilai akurasi dari algoritma Naive Bayes dan algoritma C4.5 dalam melakukan klasifikasi topik soal biologi SMA kelas 11 berdasarkan kategori topik yang telah ditentukan.
- b. Mengetahui hasil evaluasi performansi dari algoritma Naive Bayes dan algoritma C4.5 dalam melakukan klasifikasi topik soal biologi SMA kelas 11 berdasarkan kategori topik yang telah ditentukan.
- c. Membandingkan hasil akurasi dan evaluasi performansi dari algoritma Naive Bayes dan algoritma C4.5 dalam melakukan klasifikasi topik soal biologi SMA kelas 11 berdasarkan kategori topik.

I.4 Batasan Penelitian

Adapun batasan dalam melakukan penelitian ini adalah sebagai berikut:

1. Data yang digunakan semua berasal dari buku "*Siapa Pintar Belajar Mandiri*".
2. Data yang digunakan yaitu soal-soal biologi kelas 11 kurikulum 2013.
3. Berfokus pada soal yang tidak mengandung gambar dan *equation*.

I.5 Manfaat Penelitian

Manfaat penelitian ini:

1. Bagi Akademik, penelitian ini diharapkan memberikan informasi bagi peneliti atau calon peneliti untuk menerapkan ke dalam model yang lebih luas dan kompleks lagi atau sebagai bahan acuan untuk melengkapi referensi pustaka akademik.

2. Bagi penulis, penelitian ini bermanfaat untuk menerapkan ilmu yang didapat selama perkuliahan dan menambah pemahaman tentang algoritma *Naïve Bayes* dan algoritma C4.5.