

BAB I Pendahuluan

I.1 Latar Belakang

Dalam setiap proses pembelajaran guru tentunya selalu melakukan pengukuran terhadap pencapaian peserta didik. Hal tersebut dilakukan untuk mengetahui tingkat penguasaan peserta didik terhadap materi yang telah diajarkan. Penilaian merupakan komponen penting dalam sistem pendidikan untuk mengetahui perkembangan dan tingkat pencapaian hasil pembelajaran. Penilaian memerlukan data yang baik. Salah satu sumber data itu adalah hasil pengukuran. Pengukuran adalah suatu kegiatan yang ditujukan untuk mengidentifikasi besar kecilnya obyek atau gejala (Hadi, 1995). Kegiatan pengukuran ini biasanya dilakukan melalui tes. Dalam penyusunan soal, guru akan mengklasifikasikannya sesuai dengan indikator soal. Pada penentuan indikator soal yang harus diperhatikan oleh guru yaitu klasifikasi topik. Klasifikasi topik adalah proses dimana suatu system menganalisis pertanyaan dan memberi label pada pertanyaan tersebut berdasarkan kategori topiknya (Tika dan Adiwijaya, 2019). Soal pilihan ganda merupakan bentuk soal yang jawabannya dapat dipilih dari beberapa kemungkinan jawaban yang telah disediakan. Konstruksinya terdiri dari pokok soal dan pilihan jawaban. Pilihan jawaban terdiri atas kunci dan pengecoh. Kunci jawaban harus merupakan jawaban benar atau paling benar sedangkan pengecoh merupakan jawaban tidak benar, namun daya jebaknya harus berfungsi, artinya siswa memungkinkan memilihnya jika tidak menguasai materinya.

Item soal yang baik adalah item yang tingkat kesukarannya dapat diketahui tidak terlalu sukar dan tidak terlalu mudah. Arifin, 2011 mengatakan bahwa perhitungan tingkat kesukaran soal adalah pengukuran seberapa besar derajat kesukaran suatu soal. Jika suatu soal memiliki tingkat kesukaran seimbang (proporsional), maka dapat dikatakan bahwa soal tersebut baik. Tetapi, suatu soal hendaknya tidak terlalu sukar dan tidak pula terlalu mudah. Untuk memperoleh prestasi belajar yang baik, sebaiknya proporsi antara tingkat kesukaran soal tersebar secara normal (Arifin, 2011). Maka dari itu perlu dilakukan klasifikasi soal.

Salah satu konsep dari teknologi informasi yang sangat bermanfaat di bidang Pendidikan adalah *E-Learning*. *E-learning* merupakan salah satu bentuk model pembelajaran yang difasilitasi dan didukung pemanfaatan teknologi informasi dan komunikasi (Hanum, 2013). Dari segi guru, *e-learning* dapat membantu guru dalam mengelola peserta didik, memberi penugasan, diskusi, bahkan memberikan penilaian tanpa harus bertatap muka secara langsung. Hal ini membuktikan bahwa *e-learning* merupakan inovasi yang dapat dipakai sebagai alternatif pembelajaran (Abror & Jati, 2016).

Klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Ada berbagai macam metode klasifikasi, diantaranya adalah algoritma *Naive Bayes* dan algoritma *Support Vector Machine*. Algoritma *Naive Bayes* dan algoritma *Support Vector Machine* merupakan metode klasifikasi pada *text mining*.

Berdasarkan penelitian terdahulu, penelitian yang dilakukan (Izzul, Shina,& Azhar, 2020) dalam membandingkan algoritma *Naive Bayes* dan *Support Vector Machine*. Pada penelitian mereka, data yang digunakan adalah tweets yang mengandung kata “Universitas Muhammadiyah Malang”, “Muhammadiyah”, “UMM”, atau “unmuh” selama tahun 2018 hingga 2019. Proses pengumpulan data dilakukan dengan teknik *scraping* menggunakan *twitterscraper* yang bisa langsung diinstal dan digunakan melalui terminal atau cmd. Dari hasil *scraping*, didapatkan sebanyak 2654 data tweets yang disimpan ke dalam format .csv. Data yang digunakan terdiri dari 2030 *dataset* yang terbagi menjadi 749 data sentiment netral, 702 data sentimen positif dan 579 sentimen negatif. Hasil komparasi kedua metode menunjukkan bahwa *Naive Bayes* mendapatkan hasil akurasi yang lebih baik dari *SVM* dengan akurasi sebesar 73,65%.

Pada penelitian ini peneliti akan melakukan perbandingan algoritma *Naive Bayes* dan *Support Vector Machines* dalam mengklasifikasikan soal berdasarkan kategori topik pada mata pelajaran Biologi kelas XI. Algoritma *Naive Bayes* dan *Support*

Vector Machines digunakan karena dianggap cukup mudah dipahami dan telah banyak digunakan dalam studi terkait lainnya. *Naïve Bayes* memiliki kesederhanaan dalam tahap preprocessing teks dan klasifikasi teks itu sendiri. Proses klasifikasi dengan algoritma *Naïve Bayes* ini dilakukan pada saat tahap pengujian dimana *Naïve Bayes* akan mencari nilai akurasi tertinggi dari semua bank soal yang akan diujikan (Hardifa,2019). Algoritma *Support Vector Machines* dikembangkan oleh Vapnik, Guyon & Boser. Pertama kali SVM ditampilkan sekitar tahun 1992 pada Annual Workshop on Computational Learning Theory. Metode tersebut merupakan learning machine dengan tujuan mencari hyperplane terbaik yang membagi dua buah kelas di input space (Handayanto, 2019). Sehingga pada tugas akhir ini peneliti bermaksud melakukan perbandingan hasil akurasi antara algoritma *Naïve Bayes* dan algoritma SVM dalam melakukan klasifikasi topik soal pada mata pelajaran biologi SMA kelas 11 dengan menggunakan *dataset* soal biologi yang terdapat pada buku Biologi Fundamental berbasis ketrampilan abad 21 kelas 11 SMA.x`

I.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dibahas di atas, berikut pernyataan terkait rumusan masalah yang akan diidentifikasi oleh peneliti pada dokumen ini antara lain:

- a. Bagaimana klasifikasi soal biologi SMA kelas 11 berdasarkan kategori topik dengan implementasi perbandingan algoritma *Naïve Bayes* dan *Support Vector Machine*?
- b. Bagaimana cara menangani data yang *imbalance* pada studi kasus dengan menggunakan SMOTE?
- c. Bagaimana perbandingan tingkat akurasi dan evaluasi performansi dari algoritma *Naïve Bayes* dan *Support Vector Machine* dalam mengklasifikasikan soal biologi SMA kelas 11 berdasarkan kategori topik?

I.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang diuraikan diatas , maka tujuan dari penelitian ini adalah :

- a. Mengetahui perbandingan klasifikasi dengan algoritma *Naïve Bayes* dan *Support Vector Machine* untuk mengelompokkan soal pada soal biologi SMA kelas 11 berdasarkan kategori topik.
- b. Mengetahui penjelasan mengenai data *imbalance* dengan menggunakan SMOTE pada studi kasus.
- c. Membandingkan hasil akurasi dan evaluasi performansi dari algoritma *Naïve Bayes* dan algoritma *Support Vector Machine* 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. Berfokus pada perbandingan algoritma *Naïve Bayes* dan algoritma *Support Vector Machines* untuk menentukan topik soal yang akan diberikan kepada siswa.
2. Berfokus pada tipe soal yang diberikan untuk siswa yaitu pilihan ganda.
3. Data yang digunakan yaitu soal-soal biologi kelas 11.
4. Menggunakan metode SMOTE (Synthetic Minority Over-Sampling Technique) dalam menangani imbalance data atau data tidak seimbang.
5. Hasil akhir dari penelitian ini berupa perbandingan hasil akurasi yang nantinya akan diuji menggunakan *tools python*.

I.5 Manfaat Penelitian

Manfaat penelitian ini adalah sebagai berikut:

- a. Bagi Penulis, Sebagai sarana untuk menerapkan ilmu yang telah didapatkan selama perkuliahan serta menambah pemahaman dan pengalaman terutama tentang algoritma klasifikasi *Naïve Bayes* dan *Support Vector Machines* .
- b. 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 yang dapat dikembangkan bagi

pengembang konsep dan materi lebih lanjut serta dapat melengkapi referensi pustaka akademik.