

BAB I PENDAHULUAN

I.1 Latar Belakang

Di era digital saat ini pendidik harus memanfaatkan teknologi untuk menunjang proses pembelajaran. Seorang pendidik diharapkan mampu menciptakan alat penilaian dan evaluasi yang efektif. Salah satu metode yang bisa digunakan adalah dengan menyusun pertanyaan yang dapat disimpan dan digunakan kapan pun diperlukan (Kuswandi dkk., 2022). Dengan menyusun soal-soal yang terstruktur dan sesuai dengan tujuan pembelajaran, pendidik dapat memastikan bahwa penilaian yang dilakukan lebih terarah dan bermanfaat. Fakta ini sesuai dengan pendapat Octiviyani (2021) yang menyatakan bahwa analisis butir soal penting untuk menilai kualitasnya, yaitu sejauh mana soal tersebut memberikan informasi yang akurat sesuai dengan tujuan pengajaran, termasuk dalam menentukan siswa yang sudah atau belum menguasai materi yang diajarkan oleh pendidik.

Salah satu pelajaran yang umum dikenal yaitu Matematika. Matematika adalah sarana untuk memahami, mengilustrasikan pola-pola dan interaksi yang ada di dunia ini dengan menggunakan bahasa simbolis dan penalaran logis (Hayati & Jannah, 2024). Ciri khas matematika adalah penggunaan simbol-simbol untuk menyampaikan konsep dalam soal. Hal ini didukung oleh pendapat Silviana (2021) bahwa matematika dijuluki sebagai bahasa simbol karena dengan bahasa ini, siswa dapat lebih mudah mengenali dan menyelesaikan soal-soal matematika. Oleh karena itu, pendidik perlu menyusun soal-soal berdasarkan topik agar siswa dapat lebih fokus dalam memahami dan menguasai setiap konsep matematika yang diajarkan.

Selain pentingnya menyusun dan menganalisis soal, klasifikasi soal berdasarkan topik juga dibutuhkan. Mengklasifikasi soal berdasarkan topik bermanfaat bagi siswa dan guru sebab dapat membantu mereka membuat keputusan dalam memilih soal berdasarkan kategori topiknya (Muhaimin dkk., 2023). Klasifikasi adalah metode untuk mengelompokkan data berdasarkan atribut atau karakteristik yang dimiliki oleh data tersebut (Wibawa dkk., 2018). Model klasifikasi bertujuan untuk mengelompokkan materi atau konten pembelajaran yang tersedia secara gratis di internet. Model ini menggunakan teknik-teknik pembelajaran mesin

untuk mengkategorikannya ke dalam topik atau jenis tertentu (Lalitha & Sreeja, 2020). Maka dari itu, dengan adanya klasifikasi dapat membantu mengatur dan menyederhanakan akses ke sumber-sumber pendidikan, sehingga mempermudah para pembelajar dalam menemukan materi tertentu.

Namun, klasifikasi yang dilakukan secara manual akan sulit dan menghabiskan banyak waktu dibanding klasifikasi secara otomatis atau komputerisasi mengingat banyaknya data yang akan diproses (Ariandi dkk., 2023). Hal ini sejalan dengan pendapat Listiowarni & Setyaningsih (2018) bahwa proses klasifikasi data soal secara manual yang melibatkan pengecekan setiap *term* dan jumlah data yang besar menjadi sulit dan memakan waktu, oleh karena itu, diperlukan penerapan sistem otomatis yang dapat memudahkan proses ini dan menghemat waktu dengan membuatnya lebih efisien.

Berdasarkan penelitian yang telah dilakukan Permana & Siswandi (2018) yaitu klasifikasi topik soal UN Bahasa Indonesia dengan pendekatan *Stemming Porter* KBBI dan *Naïve Bayes*. Penelitian tersebut bertujuan mempermudah proses klasifikasi topik soal ujian nasional bahasa indonesia dengan komputerisasi dan mencari nilai akurasi klasifikasinya. Hasilnya, *stemming porter* KBBI menghasilkan nilai akurasi data training 93,71% sedangkan data *testing* dengan akurasi 86%. *Stemming porter* KBBI mempengaruhi hasil dari nilai akurasi klasifikasi topik soal UN dengan ketepatan kata dasar pada proses stemmingnya. Maka, dengan mengembangkan model ini dapat menjadi solusi dalam melakukan klasifikasi topik dalam bidang pembelajaran.

Dalam melakukan klasifikasi, terdapat beberapa algoritma selain *Naïve Bayes*. Algoritma lain yang dapat digunakan diantaranya, *k-Nearest Neighbor* (k-NN), *Support Vector Machine* (SVM), *Random Forest* (RF), *Decision Tree* (DT), *Decision Stump* (DS), *Rule Induction* (RI), *Linear Discriminant Analysis* (LDA), *Linear Regression* (LR), dan *Neural Network* (NN) (Saifudin, 2018). Pada penelitian ini penulis akan melakukan klasifikasi soal ujian Matematika berdasarkan topik menggunakan algoritma *Naïve Bayes*. *Naive Bayes Classifier* merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes (Wibawa dkk., 2018). *Naive bayes* merupakan salah satu teknik dalam melakukan

klasifikasi pada data *mining*. Algoritma klasifikasi *Naïve Bayes* ialah metode yang paling baik dalam membagi data *training* dan data *testing* dengan rasio yang berbeda (Andini dkk., 2021). Maka dari itu, penulis ingin mengklasifikasi soal ujian Matematika berdasarkan topik menggunakan algoritma *Naïve Bayes* dengan dataset soal ujian pilihan ganda mata pelajaran Matematika tingkat SMP yang terdapat pada bank soal. Hal ini dilakukan untuk mengurangi waktu dan upaya yang dilakukan dalam proses klasifikasi sehingga meningkatkan efektivitas dalam pengelolaan soal.

I.2 Rumusan Masalah

Berikut rumusan masalah yang akan dibahas pada penelitian ini:

1. Bagaimana performa model klasifikasi berdasarkan topik menggunakan algoritma *Naive Bayes*?
2. Bagaimana pengaruh simbol matematika terhadap hasil klasifikasi?

I.3 Tujuan Penelitian

Berdasarkan rumusan masalah diatas, maka penelitian bertujuan sebagai berikut:

1. Mengetahui seberapa optimal model dalam mengklasifikasikan soal Matematika berdasarkan kategori topik.
2. Mengidentifikasi pengaruh simbol matematika terhadap hasil klasifikasi dari penggunaan algoritma yang diterapkan.

I.4 Batasan Penelitian

Adapun batasan dalam penelitian ini ialah:

1. *Dataset* bersumber dari bank soal ujian Matematika tingkat SMP.
2. Soal yang digunakan hanya berjenis pilihan ganda.
3. Data yang digunakan hanya berbentuk teks.

I.5 Manfaat Penelitian

Manfaat dalam penelitian ini:

1. Bagi pengajar, penelitian ini diharapkan dapat mempermudah melakukan klasifikasi soal berdasarkan topik.

2. Bagi penulis, penelitian ini menjadi wadah untuk mengaplikasikan ilmu yang didapat juga menambah wawasan terlebih mengenai klasifikasi soal berdasarkan topik menggunakan algoritma *Naïve Bayes*.
3. Bagi peneliti lain, penelitian ini dapat berfungsi sebagai referensi yang bisa digunakan untuk pengembangan penelitian lebih lanjut.
4. Bagi pengembang LMS, penelitian ini dapat membantu dalam memberikan model klasifikasi bank soal.