

# BAB I PENDAHULUAN

## I.1 Latar Belakang

Pendidikan adalah salah satu komponen utama dalam meningkatkan kesejahteraan hidup manusia serta membangun sumber daya manusia yang berkualitas. Pendidikan pada era saat ini telah mengalami perubahan dan inovasi dengan mengembangkan bentuk literasi yang baru, seperti literasi digital, literasi informasi, dan literasi media (Mardhiyah dkk., 2021). Inovasi ini tidak hanya berdampak pada cara penyampaian materi, tetapi juga pada metode evaluasi pembelajaran. Salah satu komponen penting dalam proses pendidikan adalah evaluasi, yang seringkali dilakukan melalui soal-soal ujian.

Menurut Arfandy & Musdar (2020), ujian adalah metode yang digunakan untuk menilai kemampuan siswa dalam memahami dan menerapkan pengetahuan yang mereka peroleh terhadap materi yang telah diajarkan dengan menggunakan instrumen tertentu. Instrumen yang digunakan dapat berupa kumpulan soal yang harus dikerjakan atau tugas untuk menghasilkan produk tertentu. Tujuan dari proses evaluasi ini adalah untuk menentukan seberapa baik siswa telah memahami materi yang disampaikan. Meskipun demikian, pelaksanaan ujian secara konvensional seringkali menimbulkan beberapa kendala, seperti keterbatasan waktu dan ruang, serta kesulitan dalam penilaian dan pengolahan hasil ujian. Oleh karena itu, teknologi informasi hadir sebagai solusi untuk mengatasi berbagai permasalahan tersebut.

Salah satu inovasi dari teknologi informasi yang sangat bermanfaat di bidang Pendidikan adalah *E-Learning* (Muhaimin dkk., 2023). *E-learning* memiliki peranan yang sangat penting dalam era pendidikan digital saat ini, terutama selama pandemi *Covid-19* dan setelahnya. UNESCO memperkirakan bahwa lebih dari 90% siswa di seluruh dunia tidak dapat bersekolah akibat pandemi *Covid-19*, dengan lebih dari 1,5 miliar siswa terkena dampaknya (UNESCO, 2020). Pemerintah juga mengharuskan penerapan sistem Pembelajaran Jarak Jauh (PJJ) selama pandemi, yang menjadikan pandemi *Covid-19* sebagai salah satu pendorong utama peningkatan penggunaan *E-Learning* di Indonesia (Roza dkk., 2021). *E-learning* merupakan salah satu bentuk model pembelajaran yang

memanfaatkan perangkat elektronik sebagai media utamanya, serta merupakan jenis pembelajaran berbasis web yang dapat diakses melalui internet (Nuryatin, 2020). *E-Learning* memiliki kemampuan untuk otomatisasi proses penilaian dan pengolahan hasil ujian, sehingga dapat mengefesiesikan waktu dan tenaga. Salah satu bentuk model *E-learning* dalam pembelajaran adalah Pinterran dan LMS, atau singkatan dari *Learning Management System*. LMS (*Learning Management System*) adalah sistem perangkat lunak pembelajaran yang kuat dengan menggunakan teknologi web untuk meningkatkan pembelajaran siswa dengan cara yang lebih efisien (Maslov dkk., 2021).

Proses ujian secara *online* seperti dengan LMS memerlukan fasilitas mengenai pengelolaan siswa, pengajar, dan salah satunya adalah pengelolaan soal (Sidiq & Kurniadi, 2021). Oleh karena itu, dalam pengelolaan soal ujian diperlukan proses klasifikasi soal menjadi beberapa kategori sesuai dengan topiknya. Tujuannya adalah agar soal-soal tersebut menjadi sebuah soal yang terstruktur sehingga memudahkan siswa ataupun pengajar untuk mengetahui topik soal (Muhaimin dkk., 2023), seperti soal yang diujikan pada LMS. Penelitian yang dilakukan oleh Ariandi dkk., (2023), menyebutkan bahwa klasifikasi soal yang dilakukan secara manual akan sangat sulit dilakukan dan membutuhkan waktu yang lama untuk kumpulan soal yang besar dalam sebuah *database*. Klasifikasi merupakan salah satu metode yang terdapat pada *machine learning* atau pembelajaran mesin untuk mengelompokkan objek berdasarkan ciri-ciri yang dimiliki oleh objek klasifikasi (Pristyanto & Zein, 2023). Terdapat beberapa metode algoritma klasifikasi pada *machine learning*, diantaranya yaitu *K-Nearest Neighbors*, *Naïve Bayes*, *Support Vector Machine*, *Decision Tree*, *Logistic Regression* dan *Random Forest* (Sailasya & Kumari, 2021).

Namun seringkali sebuah soal memiliki lebih dari satu topik, sehingga diperlukan penyelesaian masalah klasifikasi multi-label. Dalam klasifikasi multi-label, setiap soal dapat masuk ke dalam satu atau lebih kelompok, dalam kasus ini adalah topik. Terdapat beberapa pendekatan pada implementasi klasifikasi multi-label, salah satunya adalah *Problem Transformation* (Fitriani & Yustanti, 2022). Penelitian sebelumnya mengenai klasifikasi multi-label pernah dilakukan oleh Wiraguna dkk., (2019) dengan pendekatan *Problem Transformation*. Metode *Problem*

*Transformation* dapat menyelesaikan masalah klasifikasi multi-label dengan mengubahnya menjadi masalah klasifikasi *singlelabel* dan kemudian menggabungkannya kembali hasilnya ke dalam bentuk multi-label, dengan beberapa metode seperti *Binary Relevance*, *Classifier Chain*, dan *Label Powerset* (Wiraguna dkk., 2019).

Penelitian mengenai klasifikasi multi-label teks dengan pendekatan *Problem Transformation* sudah pernah dilakukan oleh beberapa peneliti terdahulu, seperti klasifikasi multi-label pada teks hadis dengan algoritma *Random Forest* yang dilakukan oleh Wiraguna dkk., (2019) yang menemukan *Problem Transformation Binary Relevance* tanpa *stemming* memberikan performa *hamming loss* terbaik sebesar 0,0663. Serta penelitian lain pada klasifikasi multi-label teks berita dengan algoritma K-NN yang dilakukan oleh Ramadhani, (2020) menemukan bahwa metode *Binary Relevance* memiliki nilai *hamming loss* terbaik sebesar 0,11625, dan penelitian dengan teks pertanyaan Kotakode menggunakan algoritma KNN yang dilakukan oleh Fitriani & Yustanti, (2022) menemukan bahwa metode *Label Powerset* memberikan nilai *F1-Score* terbaik sebesar 87%. Namun, belum banyak penelitian yang membandingkan algoritma *Random Forest* dan *K-Nearest Neighbors* pada penerapan klasifikasi multi-label dengan pendekatan *Problem Transformation* untuk teks soal.

Oleh karena itu, penelitian ini bertujuan untuk melakukan klasifikasi multi-label pada teks soal mata pelajaran Bahasa Indonesia tingkat SMP dengan pendekatan *Problem Transformation* dan algoritma *Random Forest* serta *K-Nearest Neighbors*, dan menentukan metode *Problem Transformation* dan algoritma yang terbaik. Dengan penyesuaian dan optimasi yang tepat, metode dan algoritma ini dapat menjadi solusi yang efektif dan efisien dalam membantu proses pengelolaan soal pada *database* pembelajaran *online* terutama pada LMS. Dengan demikian, penelitian ini diharapkan dapat meningkatkan kualitas pendidikan di Indonesia.

## **I.2 Perumusan Masalah**

Rumusan masalah yang mendasari penelitian ini adalah:

- a. Model klasifikasi pada algoritma manakah yang menghasilkan performa terbaik antara *Random Forest* dan K-NN dalam klasifikasi soal multi-label?
- b. Apa metode *problem transformation* terbaik pada pengklasifikasian soal multi-label berdasarkan kategori topik dengan algoritma *Random Forest* dan K-NN?
- c. Bagaimana karakteristik dan kata kunci dari teks soal pada setiap topik?

## **I.3 Tujuan Penelitian**

Penelitian ini bertujuan untuk:

- a. Mengetahui model terbaik pada pengklasifikasian soal multi-label antara algoritma *Random Forest* dan K-NN berdasarkan kategori topik dalam studi kasus mata pelajaran Bahasa Indonesia.
- b. Mengetahui metode *problem transformation* terbaik pada pengklasifikasian soal multi-label berdasarkan kategori topik dengan algoritma *Random Forest* dan K-NN.
- c. Mengetahui karakteristik dan kata kunci pada teks soal yang dimiliki oleh setiap topik.

## **I.4 Batasan Penelitian**

Batasan yang ditetapkan pada penelitian ini sebagai berikut:

- a. Studi kasus soal yang diklasifikasikan yaitu pada mata pelajaran Bahasa Indonesia tingkat SMP.
- b. Data yang digunakan merupakan *database* soal dari Pinterran.
- c. Jenis soal yang diklasifikasikan adalah teks soal pilihan ganda tanpa gambar beserta opsi pilihan gandanya.
- d. Bahasa pemrograman yang digunakan adalah bahasa *python*.
- e. Metrik evaluasi yang digunakan untuk pengukuran adalah *F1-Score*.

## I.5 Manfaat Penelitian

Manfaat penelitian ini sebagai berikut:

1. Bagi pengajar, penelitian ini diharapkan dapat mempermudah pengajar seperti guru dan dosen dalam menggunakan ulang soal ujian pada sistem *E-learning* yang telah diklasifikasikan.
2. Bagi akademik dan kampus, penelitian ini bermanfaat dalam meningkatkan efisiensi dalam pengklasifikasian multi-label soal pilihan ganda dengan berbagai topik.
3. Bagi pengembang sistem *E-Learning* terutama LMS dan Pinterran, penelitian ini diharapkan dapat memberikan informasi dan membantu dalam memberikan model terbaik untuk klasifikasi multi-label pada bank soal.
4. Bagi peneliti lain, penelitian ini bermanfaat sebagai referensi untuk mengembangkan model klasifikasi multi-label dengan pendekatan metode *problem transformation* pada soal berdasarkan topik dengan algoritma *Random Forest* dan K-NN dalam studi kasus soal Bahasa Indonesia.