

BAB I

PENDAHULUAN

1.1 Latar Belakang

Mengikuti alur waktu zaman modern, ilmu pengetahuan sudah berkembang dengan pesat. Selain menambah pengetahuan dan wawasan, ilmu pengetahuan juga meningkatkan kualitas pendidikan terutama di Indonesia. Salah satu yang menjadi bukti adanya perkembangan tersebut yaitu peningkatan pada jumlah publikasi ilmiah berupa jurnal. Menurut data dari *Scimago Journal and Country Rank (SJR)* pada tahun 2024, Indonesia berhasil menduduki peringkat 5 di Asia dalam hal jumlah publikasi ilmiah terbanyak, dengan total perolehan 64.596 dokumen yang telah diterbitkan[1].

Artikel jurnal merupakan sumber informasi terpercaya karena memuat hasil penelitian dan melewati proses revisi yang ketat. Jurnal ilmiah umumnya dipublikasi oleh Lembaga penelitian atau perguruan tinggi, dan memiliki tingkat kepercayaan yang tinggi dalam bidang ilmu yang diteliti. Dalam dunia akademik dan riset, artikel jurnal memainkan peran yang sangat vital karena bereputasi menjadi tolak ukur keberhasilan riset peneliti, sekaligus memberi pengakuan atas kontribusi ilmiahnya dalam suatu instansi atau perguruan tinggi[2]. Karena diketahui sebagai sumber informasi primer, jurnal merupakan tempat peneliti mempublikasikan hasil riset mereka, sehingga jurnal menjadi salah satu cara paling efektif dalam mendistribusikan penemuan atau pengetahuan baru kepada komunitas ilmiah.

Sejalannya pertumbuhan publikasi jurnal yang cukup pesat, pengelolaan data jurnal menjadi kebutuhan yang semakin krusial. Lonjakan jumlah publikasi ilmiah yang didorong oleh kemajuan teknologi dan tuntutan institusi akademik telah menghasilkan kumpulan artikel dengan topik yang sangat beragam. Hal ini menghadirkan beberapa tantangan dalam hal pengorganisasian informasi, terutama ketika kumpulan artikel tidak memiliki sistem pengelompokan topik yang konsisten. Masalah ini menjadi lebih kompleks ketika kumpulan artikel yang terdiri dari beragam cakupan topik. Artikel-artikel ini kerap mempunyai struktur dan terminologi yang berbeda, sehingga sulit untuk menyaring informasi yang benar-benar selaras tanpa alat bantu yang efektif. Tanpa sistem klasifikasi yang terstruktur,

institusi dan peneliti kesulitan dalam memetakan bidang kajian dan mengelompokkan hasil riset sesuai topiknya[3]. Hal ini tidak hanya memperlambat proses pengolahan informasi, namun juga menghambat upaya dalam memahami tren keilmuan dan membuat keputusan berbasis data.

Untuk memudahkan pengorganisasian dan pemetaan bidang atau topik dalam tumpukan informasi yang ada, diperlukan sistem otomatis yang mampu mengklasifikasikan artikel jurnal berdasarkan abstrak dan kata kuncinya. Sistem ini membantu memudahkan proses pengelolaan dengan menghasilkan pengelompokan artikel berdasarkan topik secara terstruktur. Pendekatan berbasis *machine learning*, seperti algoritma *Latent Dirichlet Allocation* (LDA) dan *Support Vector Machine* (SVM) merupakan kombinasi yang menghasilkan performa tinggi dalam pengelompokan teks[4]. LDA adalah salah satu metode pemodelan topik yang populer. LDA dapat mengekstraksi topik-topik dari kumpulan dokumen dan memberikan distribusi probabilistik topik dalam setiap dokumen. Pemilihan algoritma LDA dalam penelitian ini dikarenakan LDA memiliki keunggulan dalam klusterisasi data berjumlah besar dibandingkan beberapa metode pemodelan topik lainnya[5]. Metode ini memungkinkan memberikan analisis dominan topik dalam kumpulan data artikel berdasarkan abstraknya, sehingga membantu dalam pelabelan data. Pelabelan data secara manual memerlukan waktu yang sangat lama dan dalam beberapa kasus tertentu juga membutuhkan banyak biaya, terutama jika semakin banyak data yang digunakan[6]. Maka dari itu, LDA menjadi alternatif untuk melakukan pelabelan. Namun, untuk meningkatkan akurasi klasifikasi, hasil dari LDA ini perlu diproses lebih lanjut menggunakan metode klasifikasi seperti *Support Vector Machine* (SVM), yang diketahui akan kinerja dan keandalannya dalam menangani masalah klasifikasi data[7].

Sistem informasi untuk klasifikasi artikel ilmiah berbasis *machine learning* ini memiliki potensi penerapan yang luas. Institusi seperti perpustakaan digital, kampus, atau pengelola jurnal dapat menggunakannya untuk meningkatkan efisiensi dalam pengorganisasian artikel. Selain itu, sistem ini juga dapat mendukung manajemen informasi riset secara lebih terstruktur. Penelitian ini mengusulkan perancangan sistem informasi untuk klasifikasi artikel jurnal berbasis

topic modeling dan *machine learning*, yang diimplementasikan menggunakan platform Streamlit untuk kemudahan akses dan interaksi pengguna.

Penelitian ini berkontribusi dalam meningkatkan efisiensi pengelolaan dan pemanfaatan informasi pada kumpulan artikel jurnal, sekaligus dapat mendukung dalam pengembangan pengetahuan di berbagai bidang.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Seiring dengan meningkatnya aktivitas penelitian dan publikasi di lingkungan akademik, pertumbuhan jumlah artikel ilmiah mengalami peningkatan yang sangat pesat dalam beberapa tahun terakhir. Hal ini memunculkan tantangan dalam pengelolaan informasi ilmiah, terutama dalam hal mengelompokkan dan mengorganisasi artikel secara tematis. Dengan banyaknya variasi bidang kajian dalam satu jurnal, proses pengelompokan artikel yang memanfaatkan konten aktual menjadi penting dalam membantu penyusunan repositori yang lebih terstruktur dan efisien.

Besarnya *volume* data berupa artikel yang terus bertambah, menyebabkan proses identifikasi dan pengkategorian dokumen secara manual menjadi lebih sulit karena memerlukan tenaga dan memakan waktu yang cukup lama. Hal ini juga berdampak pada pengelolaan koleksi artikel dalam jumlah besar, seperti yang terdapat pada repositori kampus atau perpustakaan digital, yang menjadi semakin kompleks. Tanpa adanya sistem yang mampu mengelompokkan artikel berdasarkan topik spesifik secara otomatis, pengelolaan informasi menjadi tidak terorganisir, sehingga menyulitkan proses manajemen data ilmiah secara efisien.

Maka dari itu, dibutuhkan sebuah solusi berupa sistem informasi yang mampu mengelompokkan artikel jurnal secara otomatis berdasarkan isi abstraknya. Solusi ini berkontribusi dalam mengoptimalkan proses pengelompokan teks agar lebih efektif, efisien, dan relevan, sehingga mendukung pengelola dalam proses pengelompokan dan pengorganisasian informasi ilmiah secara tematik, sehingga mendukung pengelolaan data jurnal yang lebih terstruktur dan sistematis.

1.2.2 Metode Usulan

Solusi yang menjadi usulan untuk penelitian ini yaitu perancangan sistem informasi klasifikasi artikel jurnal berbasis abstrak menggunakan pendekatan *machine learning*. Sistem ini dirancang untuk membantu proses pengelompokan artikel secara otomatis dan tematis, sebagai respon atas tantangan dalam pengelolaan koleksi artikel ilmiah yang terus bertambah di lingkungan akademik dan institusi riset.

Dalam implementasinya, sistem ini memanfaatkan *Topic Modeling* menggunakan algoritma *Latent Dirichlet Allocation* (LDA) untuk melakukan pelabelan data secara otomatis. LDA digunakan untuk mengekstraksi topik dalam dataset artikel jurnal dengan cara mengidentifikasi distribusi probabilitas kata dalam dokumen. Hasil dari proses ini menghasilkan topik dominan yang kemudian dimanfaatkan sebagai label kelas untuk dokumen tersebut.

Selanjutnya, label hasil LDA digunakan dalam proses klasifikasi teks menggunakan algoritma *Support Vector Machine* (SVM). SVM merupakan metode *supervised learning* yang dilatih menggunakan hasil pelabelan dari LDA. Dengan cara ini, sistem tidak hanya mampu mengelompokkan dokumen yang ada, tetapi juga dapat mengklasifikasikan dokumen baru berdasarkan konten abstraknya ke dalam topik yang sesuai.

Sistem klasifikasi ini diimplementasikan dalam bentuk aplikasi berbasis web menggunakan *platform Streamlit*, yang menyediakan antarmuka interaktif untuk memudahkan pengguna seperti peneliti, dosen, atau pengelola repositori dalam mengakses dan memanfaatkan sistem.

Dengan kombinasi metode LDA dan SVM dalam satu sistem informasi, solusi ini mengatasi permasalahan pengelompokan artikel secara manual, sekaligus meningkatkan efektivitas, efisiensi, dan relevansi dalam pengelolaan artikel ilmiah serta memperkuat infrastruktur informasi di lingkungan akademik.

1.2.3 Pertanyaan Penelitian

Berdasarkan uraian dari latar belakang, penelitian ini diselesaikan dalam rumusan masalah berikut:

1. Berapa jumlah topik optimal yang dihasilkan oleh algoritma LDA berdasarkan nilai *coherence score* terhadap artikel jurnal penelitian?

2. Berapa akurasi algoritma SVM dalam mengklasifikasikan artikel jurnal berdasarkan label topik yang dihasilkan oleh LDA?

1.3 Tujuan dan Manfaat Penelitian

1.3.1 Tujuan Penelitian

Tujuan penelitian ini berupa:

1. Menentukan jumlah topik optimal dari artikel jurnal penelitian dalam jurnal JTIK menggunakan algoritma LDA berbasis nilai *coherence score*.
2. Mengevaluasi performa algoritma SVM dalam mengelompokkan artikel jurnal ke dalam label topik yang dihasilkan oleh LDA, menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-Score*.

1.3.2 Manfaat Penelitian

Beberapa manfaat dari penelitian ini diantaranya:

1. Meningkatkan efisiensi dalam pengelompokan artikel jurnal berdasarkan kesamaan topik melalui pemrosesan otomatis terhadap abstrak.
2. Dapat membantu menyediakan solusi pelabelan otomatis yang dapat menggantikan proses pelabelan manual yang memakan waktu dan tenaga.
3. Memfasilitasi proses klasifikasi artikel baru berdasarkan distribusi topik, sehingga memudahkan institusi atau peneliti dalam menyusun repositori atau arsip digital secara lebih terstruktur.

1.4 Batasan Masalah

Agar penyusunan serta pembahasan penelitian lebih terarah, maka disusunlah batasan masalah dalam penelitian ini. Berikut merupakan batasan masalah yang digunakan :

1. *Dataset* yang digunakan dalam penelitian ini hanya berasal dari artikel Jurnal JTIK, yaitu jurnal ilmiah *peer-review* di bawah naungan Fakultas Ilmu Komputer (FILKOM), Universitas Brawijaya (UB). Selain itu, artikel jurnal yang digunakan hanya berbahasa Indonesia.
2. Ekstraksi topik hanya dilakukan pada fitur data abstrak menggunakan algoritma *Latent Dirichlet Allocation* (LDA).

3. Label pengelompokan ditentukan berdasarkan probabilitas topik tertinggi yang dihasilkan oleh LDA. Oleh karena itu, hasil pelabelan bersifat otomatis sehingga tidak dapat dipastikan memiliki ketepatan yang sama dengan pelabelan manual oleh ahli.
4. Proses klasifikasi menggunakan data abstrak sebagai fitur dan topik dominan hasil LDA sebagai label, tanpa mempertimbangkan elemen lain seperti kata kunci atau isi keseluruhan artikel.

1.5 Sistematika Penulisan

Proposal Skripsi ini disusun secara sistematis dalam 5 bab. Berikut penjelasan masing-masing bab:

BAB I : PENDAHULUAN

Bab ini menjelaskan mengenai latar belakang penelitian, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, serta sistematika dalam penulisan.

BAB II : TINJAUAN PUSTAKA

Bab ini berisi penjelasan mengenai teori-teori yang berkaitan dengan “Penerapan Topic Modeling dengan *Latent Dirichlet Allocation* (LDA) dan *Support Vector Machine* (SVM) untuk Klasifikasi Artikel Jurnal”.

BAB III : METODOLOGI PENELITIAN

Bab ini diuraikan bagaimana tahapan atau metode yang digunakan untuk mencapai tujuan dan memperoleh hasil penelitian.

BAB IV : HASIL DAN PEMBAHASAN

Bab ini menjelaskan mengenai lingkungan uji coba, tahapan pembuatan program, dan hasil uji coba.

BAB V : PENUTUP

Bab ini berisi kesimpulan dan saran berdasarkan penelitian yang telah dilakukan.