

## ABSTRAK

Pertumbuhan publikasi ilmiah yang semakin pesat menimbulkan tantangan dalam proses pencarian dan pengelompokan artikel jurnal secara manual. Proses ini memerlukan waktu dan tenaga yang besar, serta berpotensi menghasilkan ketidaksesuaian dalam pengelompokan topik. Penelitian ini mengusulkan sistem otomatisasi klasifikasi artikel jurnal berbasis abstrak dan kata kunci dengan memanfaatkan pendekatan *Latent Dirichlet Allocation* (LDA) dan *Support Vector Machine* (SVM). LDA digunakan untuk melakukan pemodelan topik dan mengekstraksi distribusi probabilitas kata dalam dokumen, yang kemudian dimanfaatkan untuk pelabelan otomatis dokumen berdasarkan topik dominan. Hasil dari proses tersebut selanjutnya digunakan sebagai fitur dalam proses klasifikasi menggunakan algoritma SVM. *Dataset* yang digunakan berupa 1.164 artikel dari jurnal JTIK yang diperoleh melalui proses *web scraping*. Hasil pemodelan topik menunjukkan bahwa jumlah topik optimal diperoleh sebanyak 9 topik pada data abstrak tanpa kata kunci dengan nilai *coherence score* sebesar 0,217, serta 7 topik pada data abstrak dengan kata kunci dengan nilai *coherence score* sebesar 0,229. Nilai pada kedua skenario tersebut termasuk dalam kategori kurang baik, yang menunjukkan bahwa topik yang dihasilkan masih kurang relevan untuk dijadikan representasi fitur meskipun memiliki keterkaitan kata yang cukup baik. Sementara itu, hasil klasifikasi menggunakan SVM menghasilkan akurasi sebesar 79% untuk klasifikasi data abstrak tanpa kata kunci dan sebesar 76% pada data abstrak disertai kata kunci. Perbedaan nilai akurasi ini menunjukkan bahwa penyertaan kata kunci meningkatkan performa klasifikasi meskipun tidak signifikan, karena kata kunci cenderung bersifat umum dan dapat mengurangi kekhususan fitur yang diekstraksi dari teks abstrak utama.

**Kata kunci :** *Latent Dirichlet Allocation, Support Vector Machine, klasifikasi artikel, topic modeling.*

## **ABSTRACT**

*The rapid growth of scientific publications presents challenges in manually searching and categorizing journal articles. This process requires substantial time and effort and may lead to inconsistencies in topic grouping. This study proposes an automated journal article classification system based on abstracts and keywords by utilizing the Latent Dirichlet Allocation (LDA) and Support Vector Machine (SVM) approaches. LDA is employed to perform topic modeling and extract probabilistic word distributions within documents, which are then used for automatic labeling based on the dominant topics. The resulting representations are subsequently used as features for the classification process using the SVM algorithm. The dataset consists of 1,164 articles from the JTIK journal obtained through web scraping. The topic modeling results show that the optimal number of topics is nine on abstract-only data with a coherence score of 0.217, and seven topics on abstract-keyword data with a coherence score of 0.229. Both scores fall within the low-quality category, indicating that the generated topics are still less relevant as feature representations, although the word associations are fairly coherent. Meanwhile, SVM classification yields an accuracy of 79% for abstracts without keywords and 76% for abstracts with keywords. This difference suggests that the inclusion of keywords improves classification performance, although not significantly, as keywords tend to be general and may reduce the specificity of features extracted from the main abstract text.*

**Keywords :** *Latent Dirichlet Allocation, Support Vector Machine, article classification, topic modeling.*