

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Dalam era saat ini, teknologi informasi semakin berkembang pesat dan memberikan kemudahan bagi masyarakat dalam mengakses berbagai informasi, termasuk publikasi ilmiah. Publikasi ilmiah sendiri merupakan informasi yang sangat diperlukan dalam aktivitas utama civitas akademika, baik untuk keperluan penelitian, penyusunan proposal, penulisan skripsi, tesis, maupun disertasi. Publikasi ilmiah cenderung menjadi referensi atau gambaran untuk menyusun sebuah karya ilmiah dan publikasi lainnya. Namun, pencarian publikasi ilmiah memiliki kesulitan tersendiri, yakni kesulitan dalam menemukan artikel yang relevan sesuai dengan bidang yang diminati. Kesulitan dalam menelusuri artikel yang sesuai menjadi masalah baru, terutama dalam menemukan artikel lain yang memiliki kemiripan konten pada jurnal akademik.

Jurnal Pendidikan Teknik Informatika dan Komputer (PINTER) merupakan jurnal yang diterbitkan oleh Program Studi Pendidikan Teknik Informatika dan Komputer Universitas Negeri Jakarta. Jurnal ini berfokus pada publikasi artikel ilmiah bidang pendidikan, ilmu komputer, sistem informasi, multimedia, jaringan komputer, rekayasa perangkat lunak, dan teknologi pendidikan. Dengan semakin banyaknya artikel yang terbit, tersebarnya konten yang beragam membuat kesulitan baru yang harus dihadapi ketika ingin menentukan kecenderungan dan tren penelitian yang ada pada Jurnal PINTER. Hingga saat ini, Jurnal PINTER telah menerbitkan lebih dari 160 artikel dari berbagai bidang. Dengan jumlah artikel yang terus bertambah, pencarian berbasis kata kunci hanya mampu mencocokkan kata yang diketik secara persis tanpa memahami maksud atau isi artikel, sehingga artikel yang relevan sulit ditemukan oleh pembaca.

Berdasarkan masalah yang terjadi, terdapat kekurangan dalam identifikasi serta pengelompokan artikel berdasarkan kemiripan konten dari publikasi ilmiah pada Jurnal PINTER. Untuk mengatasi permasalahan yang ada, diperlukan pendekatan untuk mengelompokkan artikel secara sistematis berdasarkan kemiripan konten.

Salah satu pendekatan yang dapat digunakan adalah *machine learning*, khususnya pada metode *unsupervised learning* yang tidak memerlukan data berlabel. Dalam *machine learning*, terdapat teknik *clustering* yang berfungsi untuk mengelompokkan data berdasarkan tingkat kemiripan tertentu. *Clustering* pada dokumen teks merupakan teknik yang penting untuk mengelompokkan dokumen berdasarkan kesamaan konten dan telah terbukti efektif dalam berbagai implementasi (Hasanah, 2024). Teknik *clustering* pada dokumen dapat diterapkan pada teks berbahasa Indonesia dengan hasil yang baik, sebagaimana ditunjukkan oleh penelitian (Satria dkk., 2023) yang berhasil mengelompokkan dokumen skripsi berbahasa Indonesia menggunakan algoritma *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) dengan *Silhouette Coefficient* mencapai 0,945.

Salah satu algoritma pada teknik *clustering* yang dapat digunakan untuk pengelompokan artikel adalah algoritma DBSCAN. DBSCAN merupakan algoritma *clustering* berbasis kepadatan yang memiliki kelebihan mampu mengelompokkan data tanpa harus menentukan jumlah *cluster* di awal, mampu membentuk *cluster* dengan bentuk yang tidak beraturan, serta dapat menangani *noise* atau outlier dalam data (Simbolon & Friskila, 2024). Efektivitas DBSCAN dalam text *clustering* telah dibuktikan oleh Ritonga dan Muhandhis (2021) yang menerapkan DBSCAN untuk mengelompokkan data teks berupa tweet e-commerce dan berhasil membentuk *cluster* secara otomatis tanpa menentukan jumlah *cluster* di awal, sekaligus mampu mengidentifikasi *noise* secara langsung.

Penerapan *clustering* menggunakan algoritma DBSCAN dapat dilakukan dalam menentukan kelompok artikel dari Jurnal PINTER berdasarkan kemiripan konten. Penelitian yang menggunakan teknik *clustering* dengan algoritma DBSCAN pernah dilakukan sebelumnya pada berbagai kasus di Indonesia, contohnya pada kasus kriminalitas seperti penelitian oleh Miftahurrahmi dkk. (2024) menggunakan DBSCAN untuk mengelompokkan provinsi di Indonesia berdasarkan kasus kriminalitas dengan hasil *Silhouette Coefficient* sebesar 0,68. Hasil dari penelitian-penelitian tersebut menunjukkan bahwa DBSCAN mampu menghasilkan *cluster* yang baik dan sesuai untuk data dengan karakteristik yang bervariasi. Agar kemiripan konten antar artikel dapat diukur secara semantik, representasi teks pada penelitian ini dilakukan menggunakan Word2Vec yang mampu menangkap

hubungan makna antarkata lebih baik dibandingkan pendekatan berbasis frekuensi seperti TF-IDF (Gurning dkk., 2025).

Berdasarkan uraian latar belakang masalah dan penelitian sebelumnya, meskipun Jurnal PINTER telah menerbitkan berbagai artikel ilmiah, keterbatasan masih terdapat pada belum adanya sistem yang mampu mengelompokkan artikel-artikel tersebut secara otomatis berdasarkan kemiripan konten. Oleh karena itu, perlu dilakukan penerapan algoritma DBSCAN terhadap artikel-artikel Jurnal PINTER dengan tujuan membentuk klaster yang mencerminkan kedekatan semantik antarkonten artikel yang berkaitan.

1.2 Identifikasi Masalah

Berdasarkan latar belakang masalah di atas, permasalahan yang dapat diidentifikasi yaitu:

1. Keterbatasan dalam mengidentifikasi kecenderungan pengelompokan artikel pada Jurnal PINTER berdasarkan kemiripan kontennya.
2. Belum adanya model *machine learning* untuk mengelompokkan artikel Jurnal PINTER berdasarkan kemiripan konten.
3. Kesulitan pembaca dalam menemukan artikel yang relevan karena pencarian berbasis kata kunci tidak mampu menangkap kemiripan konten secara semantik.

1.3 Pembatasan Masalah

Berdasarkan latar belakang dan identifikasi masalah yang telah dijabarkan, batasan masalah yang dapat diidentifikasi yaitu:

1. Artikel yang digunakan adalah artikel yang dipublikasikan pada Jurnal PINTER Volume 1 nomor 1 hingga Volume 8 nomor 2.
2. Fitur yang digunakan untuk pengelompokan adalah judul, abstrak, dan isi artikel.
3. Algoritma *clustering* yang digunakan adalah algoritma DBSCAN.
4. Representasi teks menggunakan metode Word2Vec
5. Evaluasi hasil *clustering* menggunakan *Silhouette Coefficient* dan *Davies-Bouldin Index*.

1.4 Rumusan Masalah

Berdasarkan latar belakang, identifikasi, serta pembatasan masalah di atas maka rumusan masalah dalam penelitian ini adalah "Bagaimana hasil performa pengklusteran artikel pada Jurnal PINTER berdasarkan kemiripan konten menggunakan algoritma DBSCAN?"

1.5 Tujuan Penelitian

Berdasarkan latar belakang, identifikasi, batasan, dan perumusan masalah di atas maka tujuan dari penelitian ini adalah untuk menganalisis hasil performa pengklusteran artikel pada Jurnal PINTER berdasarkan kemiripan konten menggunakan algoritma DBSCAN.

1.6 Manfaat Penelitian

Penelitian ini memiliki dua jenis manfaat utama, yaitu manfaat teoretis dan manfaat praktis. Kedua manfaat ini diharapkan dapat memberikan kontribusi bagi pengembangan ilmu pengetahuan serta membantu pihak-pihak yang berkepentingan dalam memahami hasil pengklusteran artikel pada Jurnal PINTER

1.6.1 Manfaat Teoritis:

1. Penelitian ini diharapkan dapat menambah kajian tentang *text mining*, khususnya penerapan algoritma DBSCAN untuk pengklusteran artikel ilmiah berbahasa Indonesia
2. Hasil penelitian ini diharapkan dapat menjadi referensi tambahan mengenai penggunaan *Silhouette Coefficient* dan *Davies Bouldin Index* untuk mengevaluasi kualitas *cluster* pada kelompok artikel jurnal.
3. Penelitian ini diharapkan dapat memperkaya literatur di bidang Pendidikan Teknik Informatika dan Komputer terkait penerapan algoritma DBSCAN dan representasi teks Word2Vec untuk pengklusteran artikel ilmiah berbahasa Indonesia pada Jurnal PINTER.

1.6.2 Manfaat Praktis:

Bagi Pengelola Jurnal PINTER:

Penelitian ini menghasilkan pengelompokan artikel Jurnal PINTER ke dalam klaster-klaster berdasarkan kemiripan konten, disertai interpretasi karakteristik setiap klaster berdasarkan kata-kata dominannya. Hasil ini dapat dimanfaatkan pengelola jurnal sebagai acuan untuk memahami sebaran konten artikel yang telah dipublikasikan, serta sebagai dasar pertimbangan dalam pengembangan sistem pengelompokan atau rekomendasi artikel di laman Jurnal PINTER pada tahap selanjutnya.

Bagi Pembaca:

Hasil pengelompokan artikel pada penelitian ini dapat membantu dosen, peneliti, dan mahasiswa dalam memperoleh gambaran awal mengenai kelompok-kelompok artikel yang terbentuk di Jurnal PINTER berdasarkan kemiripan kontennya, sehingga proses penelusuran artikel yang relevan dapat dilakukan secara lebih terarah.

