

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai analisis evolusi topik riset skripsi Informatika menggunakan *Neural Topic Modeling* (BERTopic) berbasis IndoBERT untuk *Dynamic Topic Analysis*, diperoleh kesimpulan sebagai berikut:

1. Implementasi BERTopic dengan model *embedding* IndoSBERT (denaya/indoSBERT-large) berhasil mengidentifikasi 11 topik riset yang koheren dari 473 dokumen skripsi Informatika UNSOED periode 2018–2026. Topik-topik yang teridentifikasi mencakup spektrum riset yang luas mulai dari Sistem Informasi Desa & UMKM, Deteksi & Klasifikasi Citra (CNN/*Deep Learning*), Sistem Pendukung Keputusan, hingga Analisis Sentimen & NLP dengan *Topic Coherence* (Cv) sebesar 0,5182 dan *Topic Diversity* (TD) sebesar 0,8455, menunjukkan bahwa kata kunci antar-topik koheren secara semantik sekaligus beragam secara konten.
2. Evaluasi komparatif menunjukkan bahwa BERTopic (bertopic_017) secara konsisten mengungguli LDA (lda_004) pada seluruh dimensi evaluasi: *Topic Coherence* +44,6% (0,5182 vs 0,3583), *Topic Diversity* +30,1% (0,8455 vs 0,6500), *objective score* +37,4% (0,6328 vs 0,4604), dan *composite score* +67,1% (0,8366 vs 0,5006). Keunggulan ini dikonfirmasi oleh kualitas interpretabilitas yang jauh lebih tinggi, BERTopic menghasilkan frasa bermakna seperti 'analisis sentimen' dan 'pendukung keputusan', sementara LDA menghasilkan token terstem seperti 'bas', 'kembang', 'guna' yang sulit diinterpretasikan tanpa konteks tambahan. Selain itu, distribusi topik BERTopic lebih merata (entropi 2,2436 vs 1,9439) tanpa adanya topik yang mendominasi secara berlebihan, berbeda dengan LDA yang memiliki satu topik generik menyerap 28,8% korpus.

3. *Dynamic Topic Analysis* mengungkap pola evolusi riset yang dapat dibagi menjadi tiga fase. Fase Dominasi SI Konvensional (2018–2022) ditandai oleh dominasi topik pengembangan sistem informasi berbasis web (T0, T2) yang menyerap 31,3% korpus. Fase Transisi (2022–2023) menandai pergeseran dari SI konvensional menuju topik berbasis AI. Fase Akselerasi AI (2023–2026) menunjukkan lonjakan tajam pada topik *Deep Learning* (T4: 19 dokumen di 2024), evaluasi performa *framework* (T3: 18 dokumen di 2025), dan *machine learning* (T7, T8). Klasifikasi tren menggunakan *relative slope* dengan ambang batas $\pm 0,10$ menghasilkan 5 topik *emerging* (T3, T4, T6, T7, T8 = 34,5% korpus), 3 topik *stable* (T1, T5, T10 = 30,4%), dan 3 topik *declining* (T0, T2, T9 = 35,1%).
4. *Dashboard* interaktif berbasis Laravel dan FastAPI berhasil diimplementasikan sebagai *end-to-end platform* yang mengintegrasikan seluruh *pipeline* penelitian dari akuisisi data hingga visualisasi hasil DTA dengan pemisahan peran antara Admin Jurusan (pengelola *pipeline*) dan Mahasiswa (konsumen hasil analisis dan penerima rekomendasi judul berbasis Gemini API). Pengujian *Black Box Testing* menunjukkan seluruh 11 skenario pengujian menghasilkan *output* yang sesuai dengan spesifikasi kebutuhan fungsional.

5.2 Saran

Berdasarkan keterbatasan yang ditemukan selama pelaksanaan penelitian, beberapa saran untuk pengembangan di masa mendatang adalah sebagai berikut:

1. Perluasan sumber data dan korpus. Penelitian ini terbatas pada 474 dokumen skripsi Informatika UNSOED, yang meskipun sudah memadai untuk menghasilkan 11 topik koheren, masih tergolong corpus berukuran kecil-menengah untuk *topic modeling*. Enam dari 24 skenario BERTopic mengalami klusterisasi kolaps, yang mengindikasikan batas minimum sensitivitas model pada skala *corpus* ini. Penelitian selanjutnya dapat memperluas cakupan ke lintas jurusan dalam satu universitas (misalnya membandingkan tren riset antara Informatika,

Teknik Elektro, dan Matematika) atau bahkan lintas universitas untuk menghasilkan peta tren riset yang lebih komprehensif dan komparatif pada tingkat nasional.

2. Integrasi *forecasting* tren. Penelitian ini terbatas pada identifikasi tren historis (*emerging/stable/declining*) tanpa melakukan prediksi ke depan. Pengembangan selanjutnya dapat mengintegrasikan model *time-series forecasting* untuk memprediksi topik yang berpotensi berkembang di 2–3 tahun mendatang, sehingga memberikan rekomendasi yang lebih proaktif bagi perencanaan kurikulum dan peminatan mahasiswa.
3. Integrasi dengan sistem repositori UNSOED. Saat ini, proses akuisisi data dilakukan melalui *web scraping* yang bersifat *one-time extraction* dan bergantung pada struktur HTML halaman repositori apabila tampilan situs berubah, *scraper* harus disesuaikan secara manual. Penelitian selanjutnya dapat mengusulkan integrasi langsung dengan sistem repositori digital UNSOED (repository.unsoed.ac.id) melalui mekanisme API atau *webhook*, sehingga setiap kali skripsi baru diunggah ke repositori, data secara otomatis tersinkronisasi ke dalam *dashboard* analisis topik tanpa intervensi manual.
4. Optimasi Interpretasi Topik dengan *Large Language Models* (LLM). Penelitian ini menggunakan algoritma BERTopic untuk menghasilkan representasi topik berdasarkan kata kunci utama. Namun, penamaan label topik seringkali masih memerlukan tinjauan manual agar lebih kontekstual. Pengembangan selanjutnya dapat mengintegrasikan *Generative AI* (seperti GPT-4, Llama 3, atau Gemini) untuk melakukan pelabelan otomatis (*auto-labeling*) yang lebih semantik dan deskriptif. Selain itu, penggunaan LLM dapat dimanfaatkan untuk menghasilkan ringkasan naratif mengenai sejarah dan perkembangan suatu topik dari tahun ke tahun, sehingga memudahkan pengguna dalam memahami evolusi riset tanpa harus menganalisis kata kunci secara mandiri.