

BAB I PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi yang pesat telah menghasilkan sejumlah data yang besar dalam berbagai bidang, termasuk pendidikan. Di lingkungan akademik, khususnya studi Informatika, setiap tahun menghasilkan sekitaran 1000 skripsi. Jumlah data yang besar ini menjadi tantangan tersendiri bagi pengelola dan membuat mahasiswa kesulitan mencari topik-topik penelitian yang relevan dan tren topik yang sedang berkembang [1][2]. Berdasarkan studi yang dilakukan, terdapat 60% mahasiswa mengalami kesulitan dalam menentukan topik skripsi [3].

Untuk mempermudah dalam analisis topik yang sedang berkembang, teknik topic modeling menjadi penting dalam mencari topik data yang serupa. Metode umum yang sering digunakan untuk analisis teks adalah *Latent Dirichlet Allocation (LDA)*, yang dapat mengidentifikasi topik tersembunyi dalam dokumen teks [4]. Namun, *LDA* memiliki kekurangan dalam menangkap makna kontekstual dan hanya memperkirakan kata-kata yang sering muncul dalam tiap topik [5].

Oleh karena itu, diperlukan model yang lebih canggih seperti *Bidirectional Encoder Representations from Transformers (BERT)* yang dikembangkan oleh Google untuk melengkapi kelemahan *LDA*. *BERT* memungkinkan model untuk memahami konteks setiap kata dalam sebuah kalimat secara lebih mendalam, baik yang datang sebelum maupun sesudah kata tersebut, yang sangat berguna untuk memproses teks dengan struktur bahasa alami yang kompleks [6].

Penelitian-penelitian sebelumnya telah menunjukkan keberhasilan *BERT* dalam meningkatkan penentuan topik dengan nilai coherence 2.9656, yang lebih tinggi dibandingkan *LDA* [7]. Penerapan metode ini pada dataset abstrak skripsi di Universitas Amikom Yogyakarta dapat memberikan pemahaman dan inspirasi tentang topik-topik skripsi yang relevan. Oleh karena itu, penelitian ini bertujuan untuk menentukan topik dan mengidentifikasi pola-pola topik dalam abstrak skripsi mahasiswa studi Informatika menggunakan kombinasi dua metode, yaitu *LDA* dan *BERT*.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah berikut rumusan masalah dari penelitian ini:

1. Bagaimana membangun Topic Modeling LDA Dan BERT/BERTopic pada abstrak skripsi?
2. Seberapa efektif dari metode *LDA* Dan *BERT/BERTopic* dalam menentukan topik pada abstrak Skripsi dari tahun 2020-2024?

1.3 Batasan Masalah

Berdasarkan dari latar belakang dan rumusan masalah yang telah di uraikan diatas, maka penelitian ini membatasi luasnya cakupan ruang lingkup yang akan dibahas pada skripsi ini agar tidak melebar ke topik lain Dan memudahkan penelitian dalam pengerjaan. Adapun batasan-batasan masalah sebagai berikut:

1. Penelitian terfokus pada data skripsi tahun 2020-2024 pada studi Informatika Amikom Yogyakarta.
2. Menggunakan 2 metode yaitu *LDA* dan *(BERT)/BERTopic*.
3. Data yang digunakan hanya berupa Abstract, tidak termasuk judul atau isi skripsi secara keseluruhan.
4. Menggunakan bahasa pemrograman Python Dan menggunakan Google Collab untuk melakukan running.
5. Bahasa yang digunakan dalam abtrak skripsi yang dianalisis adalah bahasa Indonesia.
6. Evaluasi model hanya menggunakan score coherence, Dan visualisasi Pyldavis untuk Metode *LDA*.
7. Evaluasi model hanya menggunakan *score coherence*, *topic word score* Dan *intertopic distance map* untuk metode *BERT/BERTopic*.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut:

1. Mengimplementasikan metode *LDA* dan *BERT/BERTopic* untuk melakukan topik modeling pada data skripsi studi Informatika.
2. Membandingkan kinerja dari kedua metode *LDA* dan *BERT/BERTopic*

untuk melakukan klusterisasi topik skripsi dalam fakultas Ilmu Komputer.

3. Mengidentifikasi topik-topik yang muncul dalam abstract skripsi mahasiswa Study Informatika berdasarkan hasil klusterisasi.

1.5 Manfaat Penelitian

Berikut manfaat penelitian ini dilakukan yaitu sebagai berikut:

1.5.1 Manfaat Segi Teoritis Dan Praktis

1. Untuk memenuhi syarat kelulusan Strata I Program studi Informatika Universitas Amikom Yogyakarta
2. Memberikan kontribusi pada literatur tentang perbandingan metode *LDA* dan *BERT* dalam konteks klusterisasi teks.
3. Membantu dalam perencanaan kurikulum dan pengembangan program studi berdasarkan tren topik penelitian yang teridentifikasi.
4. Membantu mahasiswa dalam mencari tren topik penelitian yang sedang berkembang untuk dijadikan referensi dalam pembuatan penelitian.

1.5.2 Manfaat Bagi Penelitian Selanjutnya

1. Menjadikan dasar untuk penelitian lebih lanjut dalam analisis tren penelitian menggunakan metode pembelajaran mesin dan NLP yang lebih advanced.
2. Mendorong penelitian untuk meningkatkan kinerja dari topik modeling dan efisiensi komputasi, terutama untuk dataset yang lebih besar dan kompleks.

1.6 Sistematika Penulisan

Sistematika penulisan laporan ini disusun menggunakan dasar-dasar penulisan karya ilmiah. Metode ini dilakukan agar dalam penyusunan laporan menjadi lebih terstruktur dan mudah dipahami, sistematika penulisan laporan pada skripsi ini adalah sebagai berikut:

BAB I PENDAHULUAN, Bab ini membahas tentang Latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA, Bab ini membahas tentang tinjauan pustaka, dan

dasar-dasar teori yang digunakan.

BAB III METODE PENELITIAN, bab ini didalamnya terdapat tinjauan umum tentang objek penelitian, alur penelitian, dan alat dan bahan.

BAB IV HASIL DAN PEMBAHASAN, bab ini merupakan tahapan yang penulis melakukan pengembangan penelitian, dan menjelaskan hasil yang di dapatkan selama penelitian.

BAB V PENUTUP, berisi kesimpulan dan saran yang dapat peneliti rangkum selama proses penelitian

LAMPIRAN, berisi Program github untuk membangun sebuah project.

