

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang Masalah**

Perkembangan teknologi dan ilmu pengetahuan saat ini berkembang pesat, salah satunya memanfaatkan sistem informasi repositori, yang biasa digunakan pada bidang pendidikan. Repositori adalah perpustakaan berbasis digital dimana mahasiswa dapat mengakses skripsi dan informasi civitas akademika, maupun para akademisi di luar kampus digunakan sebagai referensi penulisan karya ilmiah [1]. Repositori memiliki rancangan yang spesifik untuk meningkatkan kualitas dan mutu agar informasi dari hasil penelitian yang lebih akurat dan tepat kepada pemustaka. Saat mahasiswa akan mencari topik-topik penelitian yang terdapat di repositori guna mencari referensi penelitian, mahasiswa perlu membaca satu persatu publikasi ilmiah dari seluruh *author* (penulis). Cara tersebut tentu akan memerlukan usaha, waktu lama dan tenaga. Oleh karena itu, diperlukan suatu metode yang dapat mengotomisasi tren topik penelitian di repositori Institut Teknologi Telkom Purwokerto (ITTP).

Repositori dapat membantu pihak pustakawan dalam pendataan skripsi dan juga memudahkan mahasiswa dalam melakukan pencarian skripsi yang relevan. Langkah otomatisasi yang dapat dilakukan untuk mencari sebuah topik penelitian menggunakan pendekatan *topic modelling*. Beberapa metode untuk menentukan *topic modelling*, salah satunya *Latent Dirichlet Allocation* (LDA). LDA adalah model probabilistik generatif dan algoritma *unsupervised learning* dimana dokumen direpresentasikan sebagai campuran acak topik-topik [1]. Topik-topik tertentu dicirikan oleh distribusi kata-kata dari dokumen-dokumen yang menjadi korpus. Prinsip LDA mengklasifikasikan dokumen sesuai topik terkait. Prinsip algoritma LDA bahwa setiap kata memiliki makna semantic [1].

Sedangkan metode topik modeling lainnya yaitu *Latent Semantic Analysis* (LSA) yang juga menggunakan prinsip analisis semantic [1]. Berbeda dengan LDA, LSA mengklasifikasi dokumen menurut topik terkait. Prinsip LSA yaitu

menghitung jarak vektor untuk mengklasifikasi dokumen sesuai topik terkait. Penggunaan LSA juga jarang ditemui dibanding penggunaan LDA. Beberapa penelitian [2]–[9] menerapkan LDA untuk mencari topik dalam kumpulan korpus berbahasa Indonesia. Penelitian [2] menerapkan LDA untuk menemukan topik-topik penelitian dari korpus keperawatan yang diperoleh dari OJS jurnal SINTA. Sedangkan penelitian [8] juga menerapkan LDA untuk menemukan topik-topik penelitian para dosen di jurnal JPTEI UNY. Hasil penerapan LDA pada [2], [8] mampu mengenali jenis-jenis topik dari kumpulan korpus dokumen hasil penelitian.

Berdasarkan latar belakang diatas maka penelitian ini mengajukan tema penelitian tentang *Topic Modeling Tugas Akhir Mahasiswa Fakultas Informatika Institut Teknologi Telkom Purwokerto*. Data yang digunakan adalah abstraksi tugas akhir mahasiswa di Fakultas Informatika (FIF) sepanjang tahun 2022. Proses vektorisasi teks akan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF).

## **1.2 Perumusan Masalah**

Tren topik penelitian mahasiswa FIF dapat dikenali dari kumpulan dokumen Tugas Akhir di repositori ITTP maka bagaimana cara mengolah kumpulan dokumen tersebut menggunakan metodologi *Natural Language Processing* (NLP) yaitu metode LDA untuk menentukan tren topik secara otomatis.

## **1.3 Pertanyaan Penelitian**

Berdasarkan uraian diatas, maka peneliti merumuskan beberapa pertanyaan yang akan dibahas pada penelitian ini:

1. Bagaimana cara mengimplementasikan *topic modeling* menggunakan LDA untuk mengetahui topik-topik pada repositori ITTP?
2. Apa topik-topik yang banyak digunakan dalam penelitian Tugas Akhir mahasiswa FIF tahun 2022?

#### 1.4 Batasan Masalah

Berdasarkan rumusan masalah, tujuan penelitian dan manfaat penelitian, maka untuk mewujudkan penelitian yang sesuai dengan masalah yang ada diperoleh batasan-batasan masalah penelitian sebagai berikut:

1. Data abstraksi mahasiswa di Fakultas Informatika (FIF) pada tahun 2022 akan digunakan pada pengerjaan Tugas Akhir ini yang tersimpan dalam repositori.
2. Data hanya menggunakan Bahasa Indonesia.

#### 1.5 Tujuan Penelitian

Berdasarkan uraian tersebut maka dapat diketahui tujuan dari penelitian ini adalah untuk menerapkan TF-IDF dan LDA untuk mengetahui topik-topik pembicaraan yang sedang hangat di abstraksi mahasiswa FIF tahun 2022.

#### 1.6 Manfaat Penelitian

Berdasarkan uraian diatas maka penelitian ini mempunyai manfaat sebagai berikut:

1. Tren topik-topik yang ada pada kumpulan data dokumen teks Tugas Akhir FIF ITTP dapat dijadikan acuan mahasiswa yang hendak mencari referensi penelitian.
2. Memperkaya penelitian tentang *topic modelling*.