

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Dalam era saat ini, teknologi semakin berkembang dan semua orang sudah memiliki kemudahan dalam menemukan informasi. Perkembangan teknologi yang cukup pesat sangat membantu masyarakat terutama mahasiswa dalam menemukan informasi terkait dengan publikasi ilmiah. Publikasi ilmiah sendiri merupakan informasi yang sangat diperlukan pada aktivitas utama civitas akademika baik dalam keperluan penelitian, penyusunan proposal penelitian, penulisan skripsi, tesis maupun disertasi. Berdasarkan pada pembahasan sebelumnya, publikasi ilmiah cenderung menjadi referensi atau gambaran untuk menyusun sebuah jurnal dan publikasi. Pencarian publikasi ilmiah memiliki kesulitan tersendiri yakni, kesulitan dalam akses pencarian sebuah topik publikasi yang sesuai dengan bidang yang diminati. Kesulitan dalam mencari sebuah topik menjadi masalah baru untuk memetakan kecenderungan penelitian seorang dosen terutama pada dosen bidang komputer di berbagai universitas, salah satunya Universitas Negeri Jakarta.

Bidang komputer pada Universitas Negeri Jakarta mencakup program studi Ilmu Komputer, Pendidikan Teknik Informatika dan Komputer, Sistem dan Teknologi Informasi. Publikasi ilmiah yang dilakukan dosen pada bidang komputer di Universitas Negeri Jakarta belum bisa dikatakan memiliki kecenderungan terhadap teknologi bidang komputer secara spesifik. Berdasarkan *Association for Computing Machinery* (ACM), bidang komputer dikategorikan menjadi beberapa bidang di dalamnya yakni, *Computer Engineering, Computer Science, Information Systems, Information Technology, Software Engineering* (ACM, 2005). ACM menjadi pedoman atau panduan dalam menentukan bidang sebuah publikasi ilmiah yang dilakukan seorang dosen pada bidang komputer di Universitas Negeri Jakarta namun, tersebarnya bidang komputer di Universitas Negeri Jakarta membuat kesulitan baru yang harus dihadapi ketika ingin menentukan kecenderungan publikasi ilmiah yang dilakukan oleh dosen bidang komputer di Universitas Negeri Jakarta.

Berdasarkan masalah yang terjadi, masalah tersebut dapat dikatakan terdapat kekurangan dalam identifikasi serta pengelompokan topik dari publikasi ilmiah pada program studi bidang komputer di Universitas Negeri Jakarta. Untuk mengatasi permasalahan yang ada, diperlukan adanya penerapan *topic modeling* pada pengambilan judul publikasi ilmiah baik dalam bentuk *Paper Conference* maupun jurnal yang sudah dipublikasikan. *Topic modeling* merupakan pendekatan untuk melakukan pengelompokan data yang berupa teks berdasarkan sebuah topik tertentu dari sebuah dokumen, *topic modeling* mampu membantu dalam melakukan identifikasi yang menghasilkan tema penting dalam sebuah koleksi data tekstual judul publikasi ilmiah dosen program studi bidang komputer di Universitas Negeri Jakarta secara manual dan mampu menentukan topik dari teks dalam jumlah besar. Salah satu, metode dalam penggunaan *topic modeling* adalah *Latent Dirichlet Allocation* (LDA), LDA adalah model probabilistik generatif (Blei, 2003), dengan kelebihan mampu memahami kumpulan data yang menghasilkan data dalam bentuk probabilistik.

Penerapan *topic modeling* melalui metode LDA dapat dilakukan dalam menentukan topik dari kecenderungan publikasi ilmiah pada dosen program studi bidang komputer di Universitas Negeri Jakarta. Namun, penerapan *topic modeling* hanya akan menentukan topik – topik dari publikasi ilmiah tersebut, bukan melakukan pengelompokan topik – topik ke dalam sebuah kategori atau kluster dari topik – topik tersebut. Penelitian pemodelan menggunakan *topic modeling* dengan metode *Latent Dirichlet Allocation* (LDA) pernah dilakukan sebelumnya pada data jurnal dalam lingkup bidang kesehatan oleh (Sahria Y, 2017). Penelitian tersebut dilakukan untuk menentukan kecenderungan topik – topik pada setiap jurnal kesehatan yang telah diterbitkan, dalam pelaksanaannya terdapat pemrosesan data dari judul – judul penelitian tentang kesehatan. Hasil dari penelitian tersebut mampu menampilkan kecenderungan, salah satu topik yang dihasilkan adalah penelitian tentang status gizi.

Hasil uraian dari penelitian sebelumnya telah dikatakan berhasil untuk mengelompokkan topik- topik tertentu berdasarkan jurnal bidang kesehatan namun, untuk melakukan kluster dari data topik yang sudah ditentukan diperlukan metode *clustering* yang terdapat pada *machine learning* adalah *clustering* menggunakan

metode *K-Means Clustering*. *K-Means Clustering* merupakan metode yang mampu mengelompokkan kumpulan data dengan karakteristik yang sama ataupun dengan karakteristik yang berbeda (Agusta, 2007). Dapat diberikan kesimpulan bahwa *K-Means Clustering* adalah teknik pengelompokkan data non-hirarki yang mampu memisahkan data dengan topik yang berbeda ataupun data dengan topik yang sama ke dalam sebuah *cluster*.

Berdasarkan uraian latar belakang masalah dan penelitian sebelumnya, meskipun telah dilakukan identifikasi terhadap topik – topik publikasi ilmiah yang berdasarkan *ACM Curricula*, namun keterbatasan terdapat pada pengelompokkan topik yang dapat memunculkan kecenderungan terhadap bidang – bidang yang terdapat pada *ACM Curricula*. Sehingga, perlu dilakukan penerapan pengelompokkan melalui implementasi kluster terhadap topik yang dihasilkan oleh *Latent Dirichlet Allocation* dengan penerapan metode kluster *K-Means Clustering*, hal ini bertujuan untuk membentuk sebuah kluster sesuai dengan kecenderungan dari topik – topik yang berkaitan. Penelitian lebih lanjut butuh dilakukan untuk mengetahui topik publikasi ilmiah yang cenderung dihasilkan dari publikasi ilmiah dari para dosen program studi bidang komputer di Universitas Negeri Jakarta.

1.2 Identifikasi Masalah

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

- 1) Keterbatasan dalam mengidentifikasi topik publikasi ilmiah yang sesuai dengan bidang *ACM Curricula 2005* pada publikasi ilmiah dosen program studi bidang komputer di Universitas Negeri Jakarta yang terindeks *scopus*
- 2) Perbedaan kecenderungan topik pada penelitian dan publikasi ilmiah yang terindeks *scopus* oleh dosen program studi bidang komputer di Universitas Negeri Jakarta
- 3) Belum ada metode khusus dalam mengidentifikasi kecenderungan topik publikasi ilmiah yang terindeks *scopus* oleh dosen program studi bidang komputer di Universitas Negeri Jakarta

1.3 Pembatasan Masalah

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

- 1) Publikasi ilmiah yang diidentifikasi adalah publikasi ilmiah yang sudah terindeks *scopus* oleh dosen program studi bidang komputer di Universitas Negeri Jakarta meliputi program studi Pendidikan Teknik Informatika dan Komputer, Ilmu Komputer, dan Sistem Teknologi dan Informasi
- 2) Model dalam penelitian yang dilakukan akan menggunakan *topic modeling* dengan metode Latent Dirichlet Allocation (LDA)
- 3) Pengelompokan topik dilakukan dengan metode *K-Means Clustering* yang dievaluasi menggunakan *Davies Bouldin Index* dan *Silhouette Coefficient*

1.4 Rumusan Masalah

Berdasarkan latar belakang, identifikasi, serta pembatasan masalah di atas maka rumusan masalah dalam penelitian ini adalah “Bagaimana hasil implementasi *K-Means Clustering* pada topik yang dihasilkan oleh *topic modeling* menggunakan *Latent Dirichlet Allocation* pada prodi bidang komputer di Universitas Negeri Jakarta?”

1.5 Tujuan Penelitian

Berdasarkan latar belakang, identifikasi, batasan, dan perumusan masalah di atas maka tujuan dari penelitian ini adalah mengetahui kecenderungan topik melalui pengelompokan topik – topik dari publikasi ilmiah yang telah diidentifikasi menggunakan *topic modeling* dengan metode LDA berdasarkan *ACM Curricula 2005* melalui implementasi metode kluster *K-Means Clustering*

1.6 Manfaat Penelitian

Manfaat yang dihasilkan dari penelitian ini antara lain:

- 1) Dosen program studi bidang komputer dapat mengetahui kecenderungan topik penelitian dan publikasi ilmiah yang terindeks *scopus* berdasarkan *ACM Curricula 2005*
- 2) Dapat dijadikan sebagai acuan dalam melakukan pengembangan penelitian sesuai dengan kecenderungan topik yang sudah teridentifikasi