

BAB I

PENDAHULUAN

1.1. Latar Belakang

Salah satu *platform* terkenal yang dapat digunakan oleh para peneliti untuk mengunggah hasil dari penelitian yaitu Open Review net. Kelebihan dari Open Review net yaitu transparansi dalam proses *peer review*, dimana penulis dan pembaca dapat saling berinteraksi dengan memberikan komentar dari hasil penelitian yang sudah diunggah, penulis juga bisa memperbaiki kesalahan penelitiannya. Maka dari itu, penelitian bisa berkembang lebih baik dari sebelumnya. Open Review net memiliki beberapa kategori atau *venue*, salah satunya adalah ICLR (*The International Conference On learning Representation*) yang berkontribusi banyak penelitian di bidang kecerdasan buatan.

Jumlah publikasi di ICLR mengalami peningkatan dari tahun 2019 sampai 2023. Pada tahun 2019 ke 2020 jumlah publikasi meningkat sebesar 55.95%, lalu pada tahun 2020 ke 2021 mengalami peningkatan yang sedikit yaitu sebesar 17.22%. Tetapi, pada tahun 2021 ke 2022 mengalami kenaikan jumlah publikasi sebesar 32.91%. Selanjutnya, pada tahun 2022 ke 2023 mengalami peningkatan kembali dengan sebesar 43.76%. Dengan banyaknya penelitian yang telah dilakukan, situasi ini bisa disebut *Publication overload*. Sehingga memicu tantangan bagi para peneliti dalam menganalisis informasi yang berlebih. Akibatnya, para peneliti dapat mengalami kesulitan untuk menemukan topik utama, karena tidak mungkin juga para peneliti melakukan review semua publikasi. Maka dari itu, diperlukan identifikasi untuk popularitas topik agar para peneliti bisa menemukan topik yang dapat dikembangkan lagi. **Dengan demikian**, para peneliti dapat berkontribusi pada penelitian selanjutnya dan mengetahui fase perkembangan penelitian yang biasa dikenal dengan sebutan *state of the art*. Pada penelitian, *sate of the art* mengacu pada ilmu baru, fase perkembangan dan kemajuan dalam suatu

bidang [1]. Data yang dikumpulkan dapat mengembangkan pengetahuan yang bermanfaat memudahkan pembuatan keputusan berdasarkan fakta, mengoptimalkan mutu, dan meningkatkan inovasi [2].

Pada penelitian ini berfokus analisis tren topik yang dibahas di ICLR (*The International Conference On learning Representation*) pada tahun 2019 sampai 2023. Data dikumpulkan dan diproses menggunakan algoritma BERTopic dan LDA (*Latent Dirichlet Allocation*) dengan evaluasi *coherence value*. LDA dipilih karena metode yang efisien untuk mengidentifikasi topik yang tersembunyi dan dapat menemukan topik-topik yang berkaitan dalam sebuah data [3] dan LDA dapat menangani data yang kecil sampai data yang besar [4]. Sedangkan BERTopic dipilih karena teknik pemodelan ini termasuk baru yang diciptakan oleh Marteen Grootendorst [5], dan keunggulannya memahami kata dan konteks lebih optimal [4]. Tidak hanya itu, BERTopic lebih efektif dalam menangani data yang relatif besar [4]. Maka dari itu, **Tujuan dari penelitian ini**, yaitu menemukan topik-topik populer yang dibahas pada publikasi ICLR 2019 sampai 2023 sehingga dapat mengetahui fase perkembangan penelitian dan **dengan membandingkan hasil dari kedua metode tersebut**, dapat lebih memahami kelebihan dan kekurangan dari masing-masing metode yang digunakan untuk merepresentasikan topik pada jurnal ICLR.

Untuk mewujudkan pencapaian tersebut, data yang dianalisis adalah *abstract* dari setiap hasil penelitian pada ICLR (*The International Conference On learning Representation*). Pada penelitian ini, *abstract* dipilih sebagai fokus analisis. Karena isi *abstract* mencakup semua dari isi penelitian termasuk permasalahan, tujuan, metode yang digunakan, dan hasil tanpa harus membaca keseluruhan isi dari publikasi. Sehingga mempermudah dalam menemukan tren topik yang sering muncul pada setiap tahun.

1.2. Rumusan Masalah

Rumusan masalah dalam penelitian ini yaitu :

- a. Apa saja topik utama pada publikasi ICLR dari tahun 2019 hingga 2023 menggunakan algoritma LDA dan BERTopic. Dan bagaimana evaluasi coherence dari kedua algoritma tersebut.
- b. Bagaimana algoritma LDA dan BERTopic bisa digunakan untuk mengidentifikasi fase perkembangan penelitian tanpa melakukan review semua publikasi secara manual.

1.3. Tujuan Penelitian

Tujuan dalam penelitian ini yaitu :

- a. Menemukan topik-topik yang sering dibahas di ICLR pada tahun 2019 sampai 2023 dan membandingkan efektifitas dari algoritma LDA dan BERTopic.
- b. Membantu para peneliti mengetahui fase perkembangan penelitian dari banyaknya publikasi yang ada.

1.4. Batasan Masalah

Pada penelitian ini, terdapat batasan masalah untuk menetapkan fokus penelitian tetap sesuai dengan percobaan yang dilakukan :

- a. Penelitian memiliki fokus analisis pada *abstract* untuk menemukan topik-topik yang sering dibahas di ICLR.
- b. Dataset dikumpulkan dengan *crawling* menggunakan API dengan mendapat 14.613 data dari ICLR pada tahun 2019 sampai 2023.