

BAB I

PENDAHULUAN

1.1 Latar Belakang

Publikasi jurnal ilmiah memegang peranan dalam mendorong perkembangan ilmu pengetahuan dan teknologi. Publikasi jurnal merupakan sarana utama bagi para peneliti atau akademisi untuk membagikan pengetahuan baru dan menyebarkan temuan penelitian mereka, tanpa adanya publikasi jurnal, maka perkembangan ilmu tidak akan berkembang[1]. Pasal 4 Peraturan Menristekdikti No. 20 Tahun 2017 menetapkan bahwa setiap dosen diwajibkan untuk menghasilkan karya ilmiah yang dipublikasikan pada jurnal nasional terakreditasi maupun jurnal internasional bereputasi. Dengan mempublikasikan jurnal, terutama yang bereputasi internasional, dapat dijadikan salah satu komponen penentu kemajuan suatu negara karena hal tersebut menunjukkan bahwa negara tersebut memiliki sumber daya manusia yang unggul dan berkualitas[2].

Di era digital saat ini, pertumbuhan jumlah publikasi ilmiah semakin pesat, terutama pada jurnal-jurnal bereputasi internasional. Jurnal ilmiah yang bereputasi internasional salah satunya adalah yang terindeks di Scopus. Scopus sendiri merupakan salah satu basis data jurnal terbesar di dunia yang dipilih untuk menjadi acuan dan referensi dalam suatu penelitian[3]. Scopus memberikan penjelasan paling luas dan lengkap dari beberapa hasil riset yang diteliti di dunia. Scopus sudah menjadi salah satu sumber utama yang diterapkan oleh para akademisi, peneliti maupun institusi pendidikan tinggi untuk mengukur kualitas dari artikel yang dipublikasikan pada jurnal ilmiah. Jurnal yang terindeks scopus memiliki kualitas yang tinggi sebab mereka perlu melewati tahapan seleksi yang ketat, dengan adanya proses *peer review* atau proses peninjauan teman sejawat[4]. Bersama reputasi yang tinggi inilah, publikasi di jurnal yang terindeks scopus menjadi suatu kewajiban bagi para peneliti untuk mengoptimalkan hasil penelitian mereka.

Pada saat ini, jumlah jurnal ilmiah yang terindeks di Scopus meningkat pesat. Hal ini terjadi karena kemajuan internet memudahkan dan mempercepat proses penerbitan jurnal, baik di dalam maupun luar negeri, termasuk penerbitan di luar institusi. Sebagai contoh, menurut data dari www.elsevier.com, Elsevier

menerbitkan sekitar 420.000 artikel penelitian setiap tahunnya. Secara keseluruhan, jumlah dokumen yang telah diterbitkan hingga saat ini telah melebihi 25 juta[5]. Dengan tingginya volume data jurnal ilmiah, proses klasifikasi secara manual menjadi sangat tidak efisien. Hal ini tidak hanya memerlukan waktu yang lama, tetapi juga tingkat ketelitian yang tinggi untuk memastikan setiap jurnal diklasifikasikan ke dalam kategori bidang keilmuan yang tepat. Oleh karena itu, dibutuhkan sebuah sistem otomatis yang mampu melakukan klasifikasi secara akurat dan efisien.

Klasifikasi teks merupakan salah satu *task* dalam *Natural Language Processing* (NLP) yang bertujuan untuk mengelompokkan dokumen ke dalam satu atau lebih kategori atau label[6]. Metode ini sangat berguna dalam menangani jumlah data teks yang sangat besar dan terus bertambah setiap harinya, baik dalam bentuk artikel berita, ulasan produk, hingga dokumen ilmiah. Dengan klasifikasi teks, sistem komputer dapat secara otomatis memahami dan menyortir informasi secara lebih efisien, sehingga mempermudah pengguna dalam menelusuri dan mengambil informasi yang relevan. Secara umum, klasifikasi teks dapat dibagi menjadi dua jenis utama, yaitu *multi-class classification* dan *multi-label classification*. *Multi-class classification* adalah jenis klasifikasi di mana setiap data atau dokumen hanya dapat dikategorikan ke dalam satu kelas atau label saja, meskipun terdapat lebih dari dua kelas yang tersedia dalam sistem klasifikasi[7]. *multi-label classification* adalah jenis klasifikasi di mana setiap data atau dokumen dapat dikategorikan ke dalam lebih dari satu kelas atau label secara bersamaan berdasarkan ciri atau karakteristik data[8]. Dalam penelitian ini, diterapkan metode *multi-label classification*, karena satu jurnal dapat termasuk ke dalam lebih dari satu kategori bidang keilmuan.

Pada publikasi jurnal ilmiah, terdapat bagian penting yang disebut "*focus and scope*" yang menjelaskan ruang lingkup atau topik-topik yang ditetapkan oleh penerbit jurnal tersebut[9]. Informasi ini biasanya disediakan oleh penerbit jurnal untuk memberi gambaran kepada penulis dan pembaca mengenai bidang keilmuan apa saja yang menjadi perhatian jurnal, serta jenis penelitian seperti apa yang sesuai untuk diterbitkan. *Focus and scope* dapat mencakup deskripsi topik-topik utama, pendekatan metodologis, area aplikasi, serta tujuan akademik atau praktis dari

jurnal. Dalam konteks penelitian ini, *focus and scope* digunakan sebagai sumber utama teks untuk proses klasifikasi. Informasi yang terdapat pada bagian ini dianggap representatif terhadap bidang keilmuan jurnal, sehingga cocok dijadikan dasar untuk penentuan label dalam proses klasifikasi multi-label. Melalui pendekatan rule-based, teks *Focus and Scope* dianalisis untuk mengidentifikasi kata kunci atau frasa tertentu yang berkaitan dengan kategori keilmuan, yang kemudian digunakan untuk memberi label pada data.

Kendala yang sering dialami oleh para peneliti, dosen, maupun mahasiswa pada saat menemukan jurnal yang tepat untuk dijadikan rujukan atau tempat publikasi hasil penelitian mereka adalah menentukan apakah suatu jurnal memiliki cakupan topik atau kategori bidang keilmuan yang sesuai dengan bidang penelitian yang sedang mereka kerjakan. Proses ini umumnya masih dilakukan secara manual dengan membaca satu per satu bagian *focus and scope* dari banyak jurnal, yang tentunya memakan waktu, tenaga, dan sering kali menghasilkan kebingungan karena perbedaan istilah, bahasa, atau cakupan keilmuan yang tidak eksplisit. Akibatnya, banyak peneliti yang salah memilih jurnal target atau kehilangan kesempatan untuk mempublikasikan hasil risetnya secara tepat waktu.

Dengan keterbatasan tersebut, diperlukan sebuah solusi yang dapat membantu peneliti dalam menentukan kategori bidang keilmuan yang sesuai dengan topik penelitian mereka berdasarkan *focus and scope*. Salah satu pendekatan yang dapat digunakan adalah dengan mengklasifikasikan jurnal ke dalam berbagai label kategori bidang keilmuan berdasarkan informasi yang terkandung dalam *focus and Scope*. Dalam hal ini, metode klasifikasi multi-label menjadi solusi yang tepat, karena memungkinkan satu jurnal untuk dikategorikan ke dalam lebih dari satu kategori bidang keilmuan secara bersamaan. Untuk mendukung proses klasifikasi tersebut, digunakan pendekatan *rule-based* dalam tahap awal, di mana sistem secara otomatis mengidentifikasi dan memberikan label berdasarkan kata kunci yang terdapat dalam teks *focus and scope*. Teknik ini bekerja dengan menerapkan aturan yang dibuat untuk mengenali kata kunci yang berkaitan dengan masing-masing kategori bidang keilmuan. Setelah proses pelabelan selesai, model *Bidirectional Long Short-Term Memory* (BiLSTM) diterapkan untuk menganalisis teks secara lebih mendalam, dengan tujuan membangun sistem klasifikasi.

Metode BiLSTM dengan kemampuan *bidirectional*-nya lebih efektif memahami konteks penuh dari kata kunci jurnal tersebut. Metode BiLSTM sendiri merupakan model *LSTM* yang telah dikembangkan, karena model *LSTM* hanya memproses kata dalam satu arah yaitu *forward* (maju), sedangkan model BiLSTM dapat memproses kata dalam 2 arah yaitu *forward* (maju) dan *backward* (mundur)[10]. Keunggulan dari metode ini terletak pada tumpukan dari *LSTM* yaitu *forward layer* yang digunakan untuk menangkap informasi sebelumnya dan *backward layer* yang digunakan untuk menangkap informasi setelahnya, dari gabungan *layer* ini yang membuat BiLSTM dapat memproses informasi dengan lebih detail [11]. Dengan demikian, metode BiLSTM dapat membantu peneliti dalam menentukan jurnal yang sesuai dengan kategori bidang keilmuan yang relevan dengan topik penelitian mereka.

Beberapa penelitian sebelumnya telah menunjukkan efektivitas BiLSTM dalam klasifikasi teks, salah satunya ialah penelitian yang berjudul “*A Novel and Secured Email Classification Using Deep Neural Network with Bidirectional Long Short-Term Memory*” oleh A. Poobalan, K. Ganapriya, K. Kalaivani, K. Parthiban. Penelitian ini mengklasifikasikan email untuk mengategorikan *spam* dan *non-spam* menggunakan kombinasi metode *Deep-Neural Network (DNN)* dan *Bidirectional Long Short-Term Memory (BiLSTM)*. Hasil penelitian ini menunjukkan bahwa model DNN-BiLSTM memberikan tingkat akurasi yang lebih tinggi mencapai 98,69%, dibandingkan *Random Forest (RF)* dengan akurasi 87,93 % dan *Recurrent Neural Network (RNN)* dengan akurasi 88,69%[12].

Penelitian lain yang dilakukan oleh Rachdian Habi Yahya, Warih Maharani, Rifki Wijaya dengan judul “*Disaster Management Sentiment Analysis Using the BiLSTM Method*”. Penelitian ini mengklasifikasikan respons publik di twitter terhadap bencana alam, seperti banjir dan gempa bumi, menjadi kategori sentiment positif dan negatif. Hasil penelitian ini menggunakan metode BiLSTM dengan *Word2Vec* mencapai akurasi sebesar 84% dengan *Precision* 88%, *Recall* 92%, *F1-Score* 90%. Akurasi ini lebih baik dibandingkan metode lain yang juga diujicobakan, seperti SVM dengan akurasi 82,73% dan BERT dengan akurasi 81,78%[13].

Penelitian lain juga dilakukan oleh Akmal Perdana Hesaputra, Rayhan Digo Saputra, Yafi Hudatama Wibowo dengan judul “Identifikasi Konten Dewasa pada Cuitan Twitter Menggunakan Metode BiLSTM Sebagai Upaya Mengatasi Penyebaran Pornografi Untuk Indonesia Maju”. Penelitian ini bertujuan untuk mengklasifikasikan konten dewasa dan non-dewasa menggunakan algoritma BiLSTM *Neural Network* pada media sosial Twitter dengan memanfaatkan Twitter API dan *library Tweepy*. Hasil terbaik dari penelitian ini diperoleh dari model BiLSTM *Double layer* dengan *dropout* dengan akurasi 98,34% dan *F1-Score* 98,32%[14].

Selain itu, terdapat pula beberapa penelitian terdahulu yang mengangkat topik serupa, yaitu terkait dengan upaya mengklasifikasikan bidang keilmuan. Sebagai contoh, penelitian yang dilakukan oleh Aina Nurul Maulida, Andi Riansyah, Imam Much Ibnu Subroto dengan judul “Klasifikasi Bidang Ilmu pada Publikasi Terindeks Scopus Menggunakan Metode *K-Nearest Neighbor*” hanya menggunakan dataset dari portal SINTA, yang terbatas pada daftar publikasi nasional. Dan juga, pendekatan yang digunakan masih terbatas pada klasifikasi multi-kelas (*multi-class classification*), di mana satu dokumen hanya dikategorikan ke dalam satu dari lima kelas bidang yaitu “*Art & Humanities, Engineering & Technology, Life Science & Medicine, Natural Science, Social Science & Management.*”[15].

Penelitian lain juga dilakukan oleh Kusuma Nurnasikha, Sam Farisa Chairul Haviana, dan Imam Much Ibnu Subroto dengan judul “Klasifikasi Bidang Ilmu Publikasi Ilmiah Terindeks SINTA”. Penelitian tersebut memiliki kesamaan dengan penelitian sebelumnya, yaitu mengklasifikasikan publikasi ilmiah ke dalam bidang keilmuan, namun hanya menggunakan dataset dari portal SINTA, yang terbatas pada daftar publikasi nasional. Selain itu, pendekatan yang digunakan masih terbatas pada klasifikasi multi-kelas (*multi-class classification*), di mana setiap dokumen hanya dikategorikan ke dalam satu dari lima kelas bidang ilmu, yaitu *Art & Humanities, Engineering & Technology, Life Science & Medicine, Natural Science, dan Social Science & Management*. Input yang digunakan dalam klasifikasi adalah judul jurnal, yang kurang representatif terhadap keseluruhan konten dan cakupan keilmuan dari jurnal tersebut[16].

Berdasarkan latar belakang di atas, penelitian ini bertujuan untuk membangun model klasifikasi multi-label untuk mengklasifikasikan jurnal Scopus berdasarkan *focus and scope* menggunakan metode *Bidirectional Long Short-Term Memory* (BiLSTM) dengan teknik *Rule-Based* sebagai metode pelabelan otomatis. Melalui pendekatan ini, peneliti diharapkan dapat lebih mudah menemukan kategori bidang keilmuan yang relevan dengan topik penelitian mereka, mempercepat proses publikasi, serta mengurangi risiko kesalahan dalam menentukan jurnal target.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang, maka dapat dipaparkan rumusan masalah sebagai berikut:

1.2.1 Permasalahan

Banyak peneliti mengalami kesulitan dalam menentukan kategori bidang keilmuan yang sesuai dengan topik penelitian mereka berdasarkan *focus and scope* jurnal. Proses ini umumnya masih dilakukan secara manual dengan membaca satu per satu bagian *focus and scope* dari berbagai jurnal, yang tentunya memakan waktu, tenaga, dan sering kali menghasilkan kebingungan karena perbedaan istilah, bahasa, atau cakupan keilmuan yang tidak eksplisit. Akibatnya, banyak peneliti yang salah memilih jurnal target atau kehilangan kesempatan untuk mempublikasikan hasil risetnya secara tepat waktu.

1.2.2 Metode Usulan

Penelitian ini mengusulkan metode *Bidirectional Long Short-Term Memory* (BiLSTM) dalam mengklasifikasikan jurnal scopus secara multi-label berdasarkan kategori yang telah ditentukan menggunakan *Word2vec* sebagai ekstraksi fitur dan teknik *Rule-Based* sebagai pelabelan otomatis, sehingga setiap *focus and scope* jurnal dapat dikategorikan ke dalam satu atau lebih kategori bidang keilmuan yang sesuai secara lebih efisien.

1.2.3 Pertanyaan Penelitian

Berdasarkan latar belakang dan rumusan permasalahan yang telah dipaparkan, pertanyaan yang diajukan adalah "Berapa nilai optimal dari *hyperparameter* yang memberikan performa terbaik pada model BiLSTM melalui proses *hyperparameter tuning*?"

1.3 Tujuan dan Manfaat

1.3.1 Tujuan

Untuk mengevaluasi kinerja model klasifikasi multi-label jurnal Scopus menggunakan metode *Bidirectional Long Short-Term Memory* (BiLSTM) melalui proses *hyperparameter tuning*, dengan menggunakan metrik evaluasi seperti *Hamming Loss*, *Precision*, *Recall*, dan *F1-Score*.

1.3.2 Manfaat Penelitian

1. Memberikan informasi mengenai *focus and scope* jurnal yang telah diklasifikasikan ke dalam kategori bidang keilmuan yang relevan.
2. Menjadi referensi yang bermanfaat bagi peneliti, dosen, dan mahasiswa dalam mengidentifikasi kategori bidang keilmuan berdasarkan *focus and scope* secara cepat dan efisien.

1.4 Batasan Penelitian

Di bawah ialah batasan permasalahan yang ada pada penelitian:

1. Dataset yang digunakan dalam penelitian ini adalah jurnal-jurnal dalam bidang *Computer Science* yang tersimpan dalam database Scopus dan diambil pada bulan Oktober tahun 2024.
2. Klasifikasi multi-label dalam penelitian ini hanya menggunakan kolom *Focus and Scope* yang terdapat pada jurnal Scopus sebagai sumber utama informasi dalam menentukan kategori yang sesuai.
3. Proses pelabelan otomatis dalam penelitian ini dilakukan menggunakan teknik Rule-Based dengan menerapkan 11 aturan kategori bidang ilmu komputer. Kategori tersebut meliputi *Artificial Intelligence*, *Computational Theory and Mathematics*, *Computer Graphics and Computer-Aided Design*, *Computer Networks and Communications*, *Computer Science*, *Computer Vision and Pattern Recognition*, *Hardware and Architecture*, *Human-Computer Interaction*, *Information Systems*, *Signal Processing*, *Software*. Setiap jurnal diklasifikasikan ke dalam satu atau lebih kategori berdasarkan kecocokan kata kunci yang terdapat dalam kolom *Focus and Scope*.

4. Penelitian ini hanya memproses dan mengklasifikasikan jurnal yang memiliki *Focus and Scope* berbahasa Inggris, sehingga jurnal dengan deskripsi dalam bahasa selain Inggris belum dapat diproses dalam tahap klasifikasi.

1.5 Sistematika Penulisan

Penelitian ini dibedakan menjadi sejumlah bab, diantaranya:

- **Bab I: Pendahuluan**

Bab ini menjelaskan latar belakang, rumusan masalah, tujuan, manfaat, batasan ruang lingkup, pertanyaan penelitian, serta sistematika penulisan skripsi.

- **Bab II: Tinjauan Pustaka**

Bab ini memuat landasan teori yang mendukung penelitian, termasuk konsep-konsep utama seperti klasifikasi multi-label, arsitektur BiLSTM (*Bidirectional Long Short-Term Memory*), serta ekstraksi fitur menggunakan *Word2Vec*. Selain itu, dibahas pula penelitian terdahulu yang relevan sebagai perbandingan dan penguat dalam pengambilan metodologi.

- **Bab III: Metode Penelitian**

Bab ini menjelaskan tahapan dan metode penelitian, mulai dari pengumpulan data *Focus and Scope* jurnal Scopus, pelabelan data secara *rule-based* ke dalam 11 kategori, preprocessing teks, pembentukan vektor dengan *Word2Vec*, arsitektur model BiLSTM untuk klasifikasi multi-label, hingga evaluasi model dengan metrik yang relevan.

- **Bab IV: Hasil dan Pembahasan**

Bab ini memaparkan hasil dari eksperimen yang dilakukan, termasuk performa model BiLSTM pada berbagai konfigurasi *hyperparameter*. Disajikan pula analisis terhadap hasil evaluasi model menggunakan metrik seperti *Hamming Loss*, *Precision*, *Recall*, dan *F1-Score*.

- **Bab V: Kesimpulan dan Saran**

Berisi tentang Kesimpulan dari hasil penelitian yang telah dilakukan serta menjawab rumusan masalah yang diajukan sebelumnya. Selain itu, dalam bab ini juga disampaikan saran-saran yang diharapkan dapat bermanfaat bagi pihak-pihak terkait dan menjadi masukan untuk penelitian selanjutnya.