

# BAB I

## PENDAHULUAN

### 1.1.Latar Belakang

Scopus merupakan salah satu database jurnal internasional berkualitas tinggi dari berbagai negara yang paling banyak digunakan oleh peneliti di seluruh dunia[1]. Database jurnal tersebut dianggap sebagai salah satu basis data ilmiah terbesar di dunia yang menyediakan akses ke artikel berkualitas tinggi dari berbagai bidang ilmu yang telah diterbitkan oleh *Elsevier, Springer, Wiley, Taylor & Francis, Emerald, Nature* dll[2]. Pada database Scopus pengguna dapat mencari informasi tentang jurnal termasuk ruang dan lingkup atau *focus and scope* yang dapat dijadikan sebagai referensi dalam menemukan artikel yang cocok dengan penelitian mereka. Namun, karena banyaknya jurnal dengan berbagai cakupan *focus and scope*, sehingga diperlukan pengelompokan jurnal terindeks Scopus berdasarkan *focus and scope* dengan menggunakan klasifikasi multilabel untuk mempermudah pencarian dan pengelompokan jurnal berdasarkan topik yang relevan. Tujuan dari klasifikasi multilabel adalah untuk mengolah data teks yang memungkinkan satu data dikategorikan ke dalam lebih dari satu label secara bersamaan[3].

Metode *Long Short-Term Memory* (LSTM) menjadi pilihan yang unggul untuk mengklasifikasikan teks yang panjang dan kompleks ke dalam lebih dari satu label. LSTM adalah jenis jaringan saraf tiruan *Recurrent Neural Network* yang digunakan untuk menangani data sekuensial (berurutan) dan memiliki kemampuan untuk mengingat informasi penting selama jangka waktu yang lama[4]. Karakteristik ini menjadikan LSTM sangat relevan untuk klasifikasi multilabel, terutama pada data yang berasal dari Scopus seperti jurnal yang seringkali mencakup berbagai topik dan struktur yang bervariasi.

Efektivitas LSTM pada klasifikasi telah ditunjukkan dalam penelitian sebelumnya. Salah satunya penelitian yang dilakukan oleh Bagus Arief Hamdi Kholifatullah dan Agus Prihanto pada tahun 2023 berjudul “Penerapan Metode *Long Short-Term Memory* untuk Klasifikasi pada *Hate Speech*”. Penelitian ini menggunakan LSTM untuk mengklasifikasikan *Hate Speech* menunjukkan tingkat

akurasi data latih sebesar 86,23% dan tingkat akurasi data validasi sebesar 87,10% dengan *epoch* sebanyak 10[5].

Pada tahun 2022, Ihsan Miftahul dkk telah melakukan penelitian yang menghasilkan *LSTM* efektif untuk digunakan klasifikasi. Penelitian ini berjudul “*LSTM (Long Short-Term Memory) for Sentiment COVID-19 Vaccine Classification on Twitter*” untuk mengklasifikasikan pendapat pro(positif) dan kontra(negatif) di twitter tentang pelaksanaan vaksinasi COVID-19 yang akan dilaksanakan pemerintah. Studi ini menghasilkan berbagai proses parameter *F1-Score* sebesar 54% dan akurasi sebesar 66% pada data uji[6].

Studi mengenai klasifikasi dengan judul “Perbandingan Performa Klasifikasi Terjemahan Al-Qur'an Menggunakan Metode *Random Forest* dan *Long Short-Term Memory*” oleh Dhea Putri Aftari dkk pada tahun 2020. Hasilnya menemukan bahwa dalam kategori dakwah *LSTM* secara konsisten lebih unggul daripada *Random Forest*. *LSTM F1-Score* menghasilkan nilai sebesar 57,3% dan ketepatan 96,8%, sedangkan *Random Forest* memiliki nilai *F1-Score* sebesar 49,4% dan ketepatan 97,5%. Temuan ini menunjukkan bahwa *LSTM* memiliki kinerja yang lebih baik dalam situasi di mana penyetelan parameter yang tepat, data yang seimbang, dan *Preprocessing* yang tepat[7].

Penelitian ini menggunakan *LSTM* untuk mengklasifikasikan jurnal Scopus ke dalam lebih dari satu label sebagai *tools* pencarian ruang lingkup atau *focus and scope* yang akurat. Penelitian ini berupaya meminimalkan kesalahan klasifikasi sehingga hasil yang diperoleh dapat membantu peneliti menemukan jurnal yang sesuai dengan topik riset mereka.

## **1.2.Perumusan Masalah**

### **1.2.1. Permasalahan**

Setiap jurnal memiliki *focus and scope* yang panjang dan bervariasi, mencakup beragam topik, fokus penelitian, dan tujuan sehingga hal ini dapat memakan banyak waktu jika dilakukan peninjauan secara manual oleh pembaca. Berdasarkan situasi tersebut, maka permasalahan yang dihadapi dalam penelitian ini adalah bagaimana mengklasifikasikan multilabel jurnal Scopus berdasarkan *focus and scope*.

### 1.2.2. Metode Usulan

Penelitian ini mengajukan metode *Long Short-Term Memory* (LSTM) untuk klasifikasi multilabel berbasis *Rule-Based Labeling* pada data jurnal Scopus berdasarkan *focus and scope*.

### 1.2.3. Pertanyaan Penelitian

Berdasarkan permasalahan dan metode yang diusulkan, pertanyaan penelitian yang diajukan adalah berapa nilai *hyperparameter* yang memberikan hasil terbaik dalam proses klasifikasi multilabel jurnal Scopus ketika diuji dengan teknik pelabelan *rule-based* menggunakan metode *Long Short-Term Memory* (LSTM) melalui proses *hyperparameter tuning*?

### 1.3. Tujuan Penelitian

Berdasarkan pertanyaan penelitian, tujuan pada penelitian ini adalah menentukan nilai *hyperparameter* yang memberikan hasil terbaik dalam proses klasifikasi multilabel jurnal Scopus menggunakan teknik pelabelan *rule-based* dengan metode *Long Short-Term Memory* (LSTM) melalui proses *hyperparameter tuning*.

### 1.4. Manfaat Penelitian

1. Memberikan pemahaman tentang keterkaitan antara pengaturan nilai *hyperparameter* dan kinerja model LSTM dalam proses klasifikasi multilabel jurnal Scopus.
2. Memberikan pemahaman tentang proses klasifikasi multilabel dengan penerapan teknik pelabelan berbasis *rule-based* dalam pemodelan *Long Short-Term Memory* (LSTM).

### 1.5. Batasan Masalah

Penelitian ini dibatasi pada hal-hal berikut:

1. Menggunakan data jurnal bidang *computer science* tersimpan di database *Scopus* yang diambil pada bulan Oktober 2024 dan tidak diperbarui secara otomatis jika ada jurnal baru.

2. Klasifikasi multilabel hanya menggunakan fitur *focus and scope* yang terdapat pada jurnal Scopus sebagai sumber informasi.
3. *Rule* yang diterapkan berasal dari kategori jurnal pada Scimago.

### 1.6.Sistematika Proposal

#### 1. BAB I Pendahuluan

Bab I menguraikan tentang pentingnya klasifikasi jurnal dalam publikasi ilmiah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika proposal.

#### 2. BAB II Landasan Teori

Bab II berisi landasan teori meliputi klasifikasi teks, klasifikasi multilabel, *rule-based*, TF-IDF, *preprocessing text*, metode LSTM, dan penelitian terkait.

#### 3. BAB III Metode Usulan

Bab III berisi metode atau algoritma, arsitektur sistem, pengumpulan data, pelabelan data, *preprocessing* data, ekstraksi fitur, perhitungan manual LSTM, evaluasi, skenario pengujian, dan rencana penelitian.

#### 4. BAB IV

Bab IV menjelaskan tentang implementasi sistem termasuk implementasi pengumpulan data, implementasi pengolahan data (pelabelan *rule-based*, *preprocessing*, ekstraksi fitur TF-IDF, pembagian data), pemodelan LSTM, hasil skenari uji coba, dan deployment.

#### 5. BAB V

Bab V menjelaskan tentang kesimpulan dari hasil penelitian dan saran untuk penelitian selanjutnya.