

ABSTRAK

Pertumbuhan jumlah jurnal ilmiah yang terindeks Scopus terus mengalami peningkatan signifikan, khususnya di bidang ilmu komputer. Kondisi ini menimbulkan tantangan dalam mengklasifikasikan fokus keilmuan jurnal secara otomatis dan efisien. Penelitian ini bertujuan untuk mengembangkan model klasifikasi multi-label jurnal Scopus berdasarkan teks *Focus and Scope* menggunakan metode *Bidirectional Long Short-Term Memory* (BiLSTM). Sebelum proses pelatihan model, dilakukan pelabelan awal secara otomatis menggunakan pendekatan *rule-based* dengan memanfaatkan kata kunci dari 11 kategori bidang ilmu komputer. Selanjutnya, data melalui tahapan *preprocessing* meliputi *cleaning*, *case folding*, *tokenization*, dan *stopword removal*. Fitur teks kemudian diekstraksi menggunakan model *Word2Vec* dengan pendekatan *Skip-Gram*, sebagai representasi numerik dari kata. Data dibagi menjadi data pelatihan dan pengujian dengan rasio 80:20, lalu dilatih menggunakan arsitektur BiLSTM. Evaluasi dilakukan menggunakan metrik *Hamming Loss*, *Precision*, *Recall*, dan *F1-Score*. Hasil analisis menunjukkan bahwa kombinasi parameter paling optimal adalah *epochs* = 25 dan *batch size* = 32, dengan nilai *Precision* 0,9459, *Recall* 0,9032, *F1-Score* 0,9241 dan *Hamming Loss* terendah 0,0377. Hal ini menunjukkan bahwa model mampu mengklasifikasikan jurnal secara akurat. Dengan pendekatan ini, proses identifikasi bidang keilmuan jurnal dapat dilakukan secara lebih cepat, sistematis, dan efisien.

Kata Kunci: Klasifikasi Teks, BiLSTM, *Rule-Based*, *Word2vec*, Scopus

ABSTRACT

The number of Scopus-indexed scientific journals continues to increase significantly, especially in the field of computer science. This condition poses a challenge in classifying the scientific focus of journals automatically and efficiently. This research aims to develop a multi-label classification model of Scopus journals based on Focus and Scope text using the Bidirectional Long Short-Term Memory (BiLSTM) method. Prior to the model training process, the initial labeling is done automatically using a rule-based approach by utilizing keywords from 11 categories of computer science fields. Furthermore, the data went through preprocessing stages including cleaning, case folding, tokenization, and stopwords removal. Text features are then extracted using the Word2Vec model with the Skip-Gram approach, as a numerical representation of the word. The data is divided into training and testing data with a ratio of 80:20, and then trained using the BiLSTM architecture. Evaluation was performed using Hamming Loss, Precision, Recall, and F1-Score metrics. The analysis results show that the most optimal parameter combination is epochs = 25 and batch size = 32, with Precision value 0.9459, Recall 0.9032, F1-Score 0.9241 and the lowest Hamming Loss 0.0377. This shows that the model is able to classify journals accurately. With this approach, the process of identifying the scientific field of journals can be done more quickly, systematically, and efficiently.

Keywords: *Text Classification, BiLSTM, Rule-Based, Word2vec, Scopus*