

ABSTRAK

Scopus merupakan salah satu database jurnal internasional berkualitas tinggi dari berbagai negara yang paling banyak digunakan oleh peneliti di seluruh dunia. Pada database Scopus pengguna dapat mencari informasi tentang termasuk *focus and scope* jurnal tersebut. Namun, tiap jurnal memiliki *focus and scope* yang panjang dan bervariasi sehingga dapat memakan banyak waktu jika dilakukan peninjauan secara manual oleh pembaca. Dalam penelitian ini, menerapkan klasifikasi multilabel untuk memberikan pemahaman tentang klasifikasi jurnal dapat memiliki lebih dari satu label atau kategori. Penelitian ini menggunakan dataset berisi 1.936 jurnal pada bidang *computer science* yang diambil dari database Scopus pada Oktober 2024. Dataset ini tidak diperbarui secara otomatis ketika terdapat jurnal baru. Fitur utama yang diproses pada penelitian ini adalah *focus and scope*. Label yang digunakan terdiri dari 11 kategori bidang ilmu *computer science*, yaitu "*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*", dan "*Software*". Proses klasifikasi multilabel ini melibatkan tahapan-tahapan meliputi pelabelan data menggunakan teknik *rule-based*, *preprocessing* data, ekstraksi fitur dengan pembobotan TF-IDF, dan metode LSTM. Evaluasi model menggunakan empat metrik termasuk *micro-f1*, *hamming loss*, *precision*, dan *recall*. Nilai hyperparameter terbaik diukur berdasarkan *micro-f1* tertinggi dan *hamming loss* terendah. Pengujian hyperparameter *tuning* menunjukkan hasil terbaik pada kombinasi *epoch* 50, *batch size* 64, dan *learning rate* 0.01 memperoleh nilai *hamming loss* sebesar 0.0684, *micro-f1* 0.8471, *precision* 0.8851, dan *recall* 0.8122.

Kata kunci: Klasifikasi multilabel, *Rule-Based*, TF-IDF, LSTM

ABSTRACT

Scopus is one of the most widely used databases of high-quality international journals from various countries by researchers around the world. In the Scopus database, users can search for information about the journal's focus and scope. However, each journal has a long and varied focus and scope that can take a lot of time if the reader reviews it manually. In this study, applying multilabel classification to provide an understanding of the classification of journals can have more than one label or category. This study uses a dataset containing 1,936 journals in the field of computer science taken from the Scopus database in October 2024. This dataset is not automatically updated when there are new journals. The main features processed in this research are focus and scope. The labels used consist of 11 categories of computer science fields, namely "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 and Computer Interaction", "Information Systems", "Signal Processing", and "Software". This multilabel classification process involves stages including data labeling using rule-based techniques, data preprocessing, feature extraction with TF-IDF weighting, and the LSTM method. Model evaluation uses four metrics namely micro-f1, hamming loss, precision, and recall. The best hyperparameter value is measured based on the highest micro-f1 and lowest hamming loss. Hyperparameter tuning testing shows the best results in the combination of epoch 50, batch size 64, and learning rate 0.01 obtaining a hamming loss value of 0.0684, micro-f1 0.8471, precision 0.8851, and recall 0.8122.

Keywords: *Multilabel classification, Rule-Based, TF-IDF, LSTM.*