ABSTRAK

Meningkatnya kompleksitas sistem komputer membuat deteksi anomali pada log file menjadi krusial untuk menjaga keamanan dan keandalan. Namun, tugas ini dihambat oleh volume data yang besar dan format yang tidak terstruktur. Penelitian ini mengusulkan implementasi model Bidirectional Auto-Regressive Transformers (BART), sebuah arsitektur transformer yang menggabungkan encoder dua arah dan decoder auto-regresif, yang potensinya belum pernah diuji untuk tugas pendeteksian anomali log. Penelitian ini menggunakan metode fine-tuning pada model BART untuk tugas klasifikasi sekuens (sequence classification) dengan menggunakan dataset publik HDFS, BGL, dan Thunderbird. Model yang dihasilkan juga diterapkan dalam rancangan prototipe aplikasi untuk mengunggah log dan menampilkan hasil deteksi anomali. Hasil evaluasi menunjukkan performa yang tinggi, dengan F1-Score sebesar 0,997 pada dataset HDFS, 0,989 pada BGL, dan 0,980 pada Thunderbird, terutama pada skenario pembagian data 70:30. Berdasarkan perbandingan referensial dengan LogBERT dan LogGPT, BART memperoleh F1-Score lebih tinggi pada dataset HDFS dan BGL serta menunjukkan performa yang kompetitif pada Thunderbird. Hasil tersebut menunjukkan bahwa BART mampu mempelajari pola kontekstual pada sequence log dan mendeteksi anomali dengan tingkat presisi yang tinggi.

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG

Kata Kunci: Deteksi Anomali, Log File Server, Natural Language Processing (NLP), BART, Transformer, Deep Learning

ABSTRACT

The increasing complexity of computer systems makes anomaly detection in log files crucial for maintaining system security and reliability. However, this task is challenged by the large volume and unstructured format of log data. This study implements Bidirectional and Auto-Regressive Transformers (BART), a Transformer architecture that combines a bidirectional encoder with an autoregressive decoder, whose potential has not previously been evaluated for log anomaly detection. The BART model is fine-tuned for a sequence classification task using the publicly available HDFS, BGL, and Thunderbird datasets. The resulting model is also incorporated into a prototype application designed to upload log files and display anomaly detection results. The evaluation results demonstrate high performance, with F1-Scores of 0.997 on HDFS, 0.989 on BGL, and 0.980 on Thunderbird, particularly under the 70:30 data-splitting scenario. Based on a referential comparison with LogBERT and LogGPT, BART achieves higher F1-Scores on the HDFS and BGL datasets and remains competitive on Thunderbird. These results indicate that BART is effective in capturing contextual patterns in log sequences and detecting anomalies with high precision.

Keywords: *Anomaly Detection, Server Log Files, Natural Language Processing (NLP), BART, Transformer, Deep Learning*

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG