

ABSTRAK**“KLASIFIKASI ARTIKEL PADA CITATION NETWORK MENGGUNAKAN
GRAPH CONVOLUTIONAL NETWORK (GCN)”**

Studi Kasus pada Dataset DBLP

Tahun 1990 - 2020

(Aghata Yoan Cherly)

Universitas Sanata Dharma

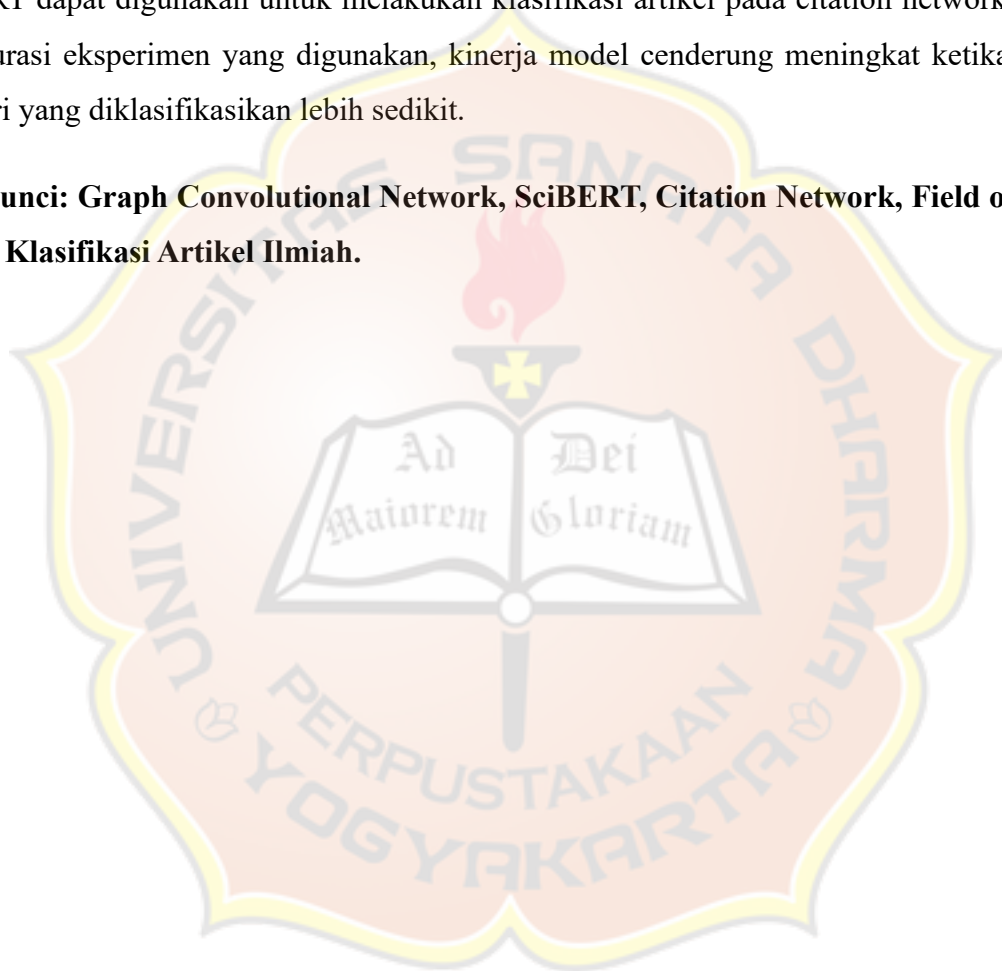
2026

Pertumbuhan jumlah publikasi ilmiah yang terus meningkat menimbulkan kebutuhan akan metode klasifikasi artikel yang mampu memanfaatkan tidak hanya informasi tekstual, tetapi juga hubungan antarartikel. Penelitian ini bertujuan untuk mengevaluasi kinerja Graph Convolutional Network (GCN) dalam melakukan klasifikasi artikel ilmiah pada citation network berdasarkan kategori Field of Study (FoS).

Dataset yang digunakan adalah DBLP Citation Network V12 dengan publikasi pada rentang tahun 1990-2020. Data melalui tahap prapemrosesan untuk memperoleh artikel yang memiliki judul, abstrak, FoS, dan referensi sitasi. Judul dan abstrak setiap artikel kemudian diproses menggunakan SciBERT untuk menghasilkan representasi fitur node berdimensi 768. Artikel direpresentasikan sebagai node, sedangkan hubungan sitasi antarartikel direpresentasikan sebagai edge. Model GCN yang digunakan terdiri atas dua lapisan dengan dimensi tersembunyi sebesar 256 dan dropout 0,5. Proses pelatihan menggunakan optimizer Adam dengan learning rate 0,01 selama lima epoch. Data berlabel dibagi menjadi 80% data pelatihan, 10% data validasi, dan 10% data pengujian. Data validasi digunakan untuk menentukan model terbaik berdasarkan Macro F1-Score, sedangkan data pengujian digunakan untuk mengevaluasi performa akhir model. Pengujian dilakukan pada tiga skenario klasifikasi, yaitu Top-20, Top-10, dan Top-5 kategori FoS, dengan metrik evaluasi berupa Accuracy, Macro Precision, Macro Recall, dan Macro F1-Score.

Hasil pengujian menunjukkan bahwa skenario Top-20 memperoleh Accuracy sebesar 49,48%, Macro Precision sebesar 41,65%, Macro Recall sebesar 31,10%, dan Macro F1-Score sebesar 30,61%. Pada skenario Top-10, diperoleh Accuracy sebesar 59,98%, Macro Precision sebesar 51,18%, Macro Recall sebesar 48,86%, dan Macro F1-Score sebesar 47,25%. Sementara itu, skenario Top-5 menghasilkan kinerja terbaik dengan Accuracy sebesar 77,27%, Macro Precision sebesar 76,33%, Macro Recall sebesar 76,73%, dan Macro F1-Score sebesar 76,39%. Hasil tersebut menunjukkan bahwa GCN dengan fitur tekstual yang dihasilkan SciBERT dapat digunakan untuk melakukan klasifikasi artikel pada citation network. Dalam konfigurasi eksperimen yang digunakan, kinerja model cenderung meningkat ketika jumlah kategori yang diklasifikasikan lebih sedikit.

Kata kunci: Graph Convolutional Network, SciBERT, Citation Network, Field of Study, DBLP, Klasifikasi Artikel Ilmiah.



ABSTRACT***“ARTICLE CLASSIFICATION ON CITATION NETWORKS USING GRAPH
CONVOLUTIONAL NETWORK (GCN)”***

Case Study on the DBLP Dataset

(1990-2020)

(Aghata Yoan Cherly)

Sanata Dharma University

(2026)

The continuous growth of scientific publications has created a need for article classification methods that can utilize not only textual information but also relationships among articles. This study aims to evaluate the performance of a Graph Convolutional Network (GCN) for classifying scientific articles in a citation network based on Field of Study (FoS) categories.

The DBLP Citation Network V12 dataset was used in this study, covering publications from 1990 to 2020. The data were preprocessed to obtain articles containing titles, abstracts, FoS information, and citation references. The title and abstract of each article were then processed using SciBERT to generate a 768-dimensional node feature representation. Articles were represented as nodes, while citation relationships between articles were represented as edges. The GCN model consisted of two layers with a hidden dimension of 256 and a dropout rate of 0.5. The model was trained for five epochs using the Adam optimizer with a learning rate of 0.01. The labeled data were divided into 80% training data, 10% validation data, and 10% testing data. The validation data were used to select the best model based on the Macro F1-Score, while the testing data were used to evaluate the final performance of the selected model. Three classification scenarios were evaluated, namely Top-20, Top-10, and Top-5 FoS categories, using Accuracy, Macro Precision, Macro Recall, and Macro F1-Score as evaluation metrics.

The experimental results show that the Top-20 scenario achieved an Accuracy of 49.48%, Macro Precision of 41.65%, Macro Recall of 31.10%, and Macro F1-Score of 30.61%. The Top-10 scenario achieved an Accuracy of 59.98%, Macro Precision of 51.18%, Macro Recall of 48.86%, and Macro F1-Score of 47.25%. Meanwhile, the Top-5 scenario achieved the best performance, with an Accuracy of 77.27%, Macro Precision of 76.33%, Macro Recall of 76.73%, and Macro F1-Score of 76.39%. These results indicate that GCN with textual features generated by SciBERT can be used to classify scientific articles in a citation network. Under the experimental configuration used in this study, the model tended to achieve better classification performance when fewer categories were included.

Keywords: Graph Convolutional Network, SciBERT, Citation Network, Field of Study, DBLP, Scientific Article Classification.

