

## Abstrak

Klasifikasi multi label untuk menentukan kategori buku perlu dilakukan untuk mempermudah pengunjung perpustakaan UIN Sunan Gunung Djati Bandung dalam menemukan buku dengan jurusan yang sesuai. Dilansir dari *website* perpustakaan UIN Sunan Gunung Djati Bandung, kata kunci pencarian terpopuler selama satu tahun terakhir menunjukkan kata kunci untuk jurusan yang sesuai. Penelitian ini bertujuan untuk membangun model yang mampu mengklasifikasikan buku dengan teknologi pembelajaran mesin, khususnya klasifikasi teks. Metode yang digunakan adalah *K-Nearest Neighbor* (KNN) dengan *metric cosine similarity* dan tambahan *grid search cv* untuk menentukan parameter yang optimal. Hasil penelitian menunjukkan bahwa kombinasi KNN, *cosine similarity* dapat diterapkan pada model untuk klasifikasi multi-label dengan beberapa skenario nilai *K*. Tingkat akurasi tertinggi model dengan data *splitting* 80% data *training* 20% data *testing* dan parameter *K* = 5 mencapai 75%, *hamming loss* sebesar 0.05%, *precision* 87%, *recall* 79%, *f1-score* 83%. Keterbatasan dataset serta distribusi data yang tidak seimbang dapat memengaruhi performa model klasifikasi. Pada penelitian ini, parameter nilai *K* berpengaruh dalam tingkat akurasi model namun tidak signifikan. Namun pemilihan parameter yang sesuai dapat membuat performa model menjadi optimal.

Kata kunci: KNN, Multi Label, Klasifikasi Buku, *Grid Search CV*



## **Abstract**

*A multi-label classification to determine book categories is essential to facilitate visitors at the UIN Sunan Gunung Djati Bandung library in finding books relevant to their respective fields. According to the UIN Sunan Gunung Djati Bandung library website, the most popular search keywords over the past year indicate a high demand for books categorized by academic disciplines. This research aims to develop a model capable of classifying books using machine learning technology, specifically text classification. The method used is K-Nearest Neighbor (KNN) with cosine similarity metric, enhanced by grid search cross-validation to determine optimal parameters. The study results show that the combination of KNN and cosine similarity can be applied to multi-label classification models under several K-value scenarios. The highest accuracy rate of the model, with an 80% training data and 20% testing data split and a parameter  $K = 5$ , reached 75%, with a hamming loss of 0.05%, precision of 87%, recall of 79%, and F1-score of 83%. The limitations of the dataset and the imbalanced data distribution may affect the performance of the classification model. In this study, the K-value parameter influenced the model's accuracy rate, though not significantly. However, selecting the appropriate parameters can optimize model performance.*

*Keywords: KNN, Multi Label, Book Classification, Grid Search CV*

