



PROJECT REPORT

Topic-Based Clustering and Social Network Analysis of News Headlines Using LDA and DBSCAN

**NORLI
22.K1.0054**

**Faculty of Computer Science
Soegijapranata Catholic University
2026**

ABSTRACT

Online news is growing rapidly every day, making it difficult for readers to find reading materials that suit their interests and understand issues currently being discussed in society. To address these issues, an automatic clustering method is needed to assist the information analysis process. The purpose of this study is to apply Latent Dirichlet Allocation (LDA) in extracting topic representations from news headlines and evaluate the effectiveness of the topic distribution in the clustering process using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method. The data used are a collection of news headlines from three online media, namely detik.com, kompas.com, and CNNIndonesia. The research stages include data pre-processing, text weighting, vector representation, application of LDA for topic formation, and application of DBSCAN to form clusters based on data density. Parameter determination is carried out through experiments to obtain the optimal number of topics in LDA as well as the appropriate epsilon and MinPts values in DBSCAN. The quality of the topic model is evaluated using the coherence score, while the quality of the cluster is assessed using the silhouette score. Furthermore, Social Network Analysis (SNA) was used to analyze the relationship structure between news headlines within each cluster through centrality and density measurements. The results indicated that the topic representation generated by LDA was effective as a basis for density-based clustering, and the network analysis was able to provide a deeper understanding of the relationship structure and the central role of news headlines within each cluster.

Keyword: LDA, DBSCAN, SNA, preprocessing, online news clustering