Classification of Sundanese Script with Classical Machine Learning

Khairat Hayati[1], Dr.Mahmud Dwi Sulistiyo,S.T.,M.T.[2],Dr.Aditya Fiman Ihsan,S.Si.M.Si.[3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]khairathayati@students.telkomuniversity.ac.id, [2]mahmuddwis@telkomuniversity.ac.id,
[3]adityaihsan@telkomuniversity.ac.id

Abstract
This research presents Sundanese script image pattern recognition through feature extraction using Histogram of Oriented Gradient (HOG). Despite the rapid development of modern technology, classification methods using classical machine learning are still relevant due to their reliability in character recognition, including Sundanese script. This study compares several classical machine learning methods such as SVM, k-NN, Decision Tree, Naïve Bayes, Random Forest, and ANN in Optical Character Recognition (OCR) for Sundanese characters. Experimental results show the superiority of Random Forest with accuracy, precision, recall, and F1-score reaching 97.02%, 97.05%, 97.01%, and 97.01% respectively, surpassing the performance of KNN, Decision Tree, Naïve Bayes, SVM, and ANN. This study aims to identify the best classical machine learning method for Sundanese script recognition through OCR, confirming that Random Forest remains effective. It also improves the understanding of the performance of various classical machine learning methods in character recognition, supporting the development of more advanced Sundanese script recognition technology in the future.

Translated with DeepL.com (free version)
Keywords: *Sundanese script, Image, Classification, Machine Learning, KNN, SVM, , Decision Tree, ANN, Random Forest, Naïve Bayes, OCR, Feature Extraction*