

Working with the standard ML pipeline in WEKA is an informative experience to understand the importance of feature engineering and model evaluation in Machine Learning. After manually extracting features like length, gender, vowel_count, and starts_with_vowel from the name column, I trained a J48 decision tree model using 10-fold cross-validation. The model achieved an accuracy of 86%, with a Kappa statistic of 0.72, indicating moderate agreement between the predicted and actual classes. The pruned tree is simple, with two leaves, effectively classifying names based on whether they started with a vowel. The results showed balanced precision and recall for both classes, demonstrating how the pipeline effectively connects feature extraction, model training, and evaluation.