ABSTRACT

APPLICATION OF TEXT MINING FOR CLASSIFICATION OF ITE LAW ARTICLE TYPES USING THE K-NN ALGORITHM

By
Alif Alpian Sahrul Muharom
NIM 20102007

Digital transformation has improved efficiency and accessibility but has also created challenges such as a data explosion and risks to individual privacy due to cybercrime. This study examines the classification of violations under the Electronic Information and Transactions Law (UU ITE) using the K-Nearest Neighbor (K-NN) algorithm. The data includes 324 cases of violations, focusing on Articles 27 and 28. Issues such as personal data protection, defamation, and online gambling have become increasingly urgent due to threats like hacking and data breaches. A text-based approach to classifying UU ITE violations can support legal efforts in the digital domain. This research addresses the need for more accurate methods of categorizing violations based on case chronology. It involves text preprocessing, weighting, and modeling using K-NN. To address data imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was employed. Dimensionality reduction was performed using Principal Component Analysis (PCA) to enhance efficiency. The initial model achieved 59% accuracy. After applying SMOTE and PCA, accuracy increased to 82%. The model performed best on Article 27 Paragraph 2 and worst on Article 27 Paragraph 1, with an average macro F1-score of 0.82, indicating balanced performance. This study demonstrates the potential of text mining to support digital law enforcement.

Keywords: *K-Nearest Neighbor (K-NN), Text Mining, Digital transformation, UU ITE*