2025 International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS)

979-8-3315-1332-0

***Abstract*—Multi-label classification is a critical task in text analysis, particularly for complex datasets like Qur'an verses, which often encapsulate multiple thematic labels. This study investigates the use of ensemble methods by combining traditional machine learning models, such as Support Vector Machine (SVM) and Naïve Bayes, with the transformer-based BERT model. The research evaluates individual and ensemble performances under varying preprocessing conditions and uses Hamming Loss as the primary evaluation metric. SVM emerged as the most effective standalone model, achieving the lowest Hamming Loss of 0.0881, while the SVM and Naïve Bayes ensemble demonstrated competitive results with a Hamming Loss of 0.0891. Interestingly, minimal preprocessing outperformed extensive text transformations, underscoring the importance of preserving semantic richness in Qur'an verses analysis. The inclusion of BERT in ensembles, while promising, often underperformed due to its sensitivity to small datasets and contextual dependencies. This research highlights the potential of integrating traditional machine learning and transformer-based models for Qur'an text classification and provides valuable insights for optimizing multi-label classification strategies in similar contexts.**

Keywords—Multi-label classification, ensemble methods, BERT, Qur'an, SVM, Naïve Bayes