979-8-3315-1332-0

I. INTRODUCTION

The Qur'an, a fundamental guide for Muslims, comprises 114 chapters and 6,236 verses addressing a wide array of topics, including faith, worship, ethics, laws, and history [1]. Each verse often encapsulates multiple themes, making the Qur'an a natural candidate for multi-label classification, where a single text can belong to several categories. For instance, verse 8 of Surah Al-Baqarah highlights both faith and science, exemplifying this complexity [2].

Multi-label classification plays a crucial role in organizing and analyzing intricate datasets like Qur'an verses [3]. Previous research has employed traditional machine learning methods, such as Naïve Bayes and its advanced variant, Tree-Augmented Naïve Bayes (TAN), achieving Hamming Loss values of 0.1247 and 0.1121, respectively [4]. While neural network-based approaches have also been applied, their computational demands and complexity present challenges for practical implementation [5].

Ensemble methods, which combine multiple models to improve classification accuracy, provide a promising alternative for addressing the complexities of Qur'an verse classification [6]. These methods leverage the strengths of individual classifiers, such as the statistical rigor of traditional machine learning models and the contextual depth of transformer-based architectures, to enhance overall performance [7][8]. However, the integration of modern transformers like Bidirectional Encoder Representations from Transformers (BERT) with traditional models remains underexplored in the domain of Qur'an verses analysis.

This research seeks to fill this gap by evaluating single classifiers Support Vector Machines (SVM), Naïve Bayes, and BERT alongside ensemble methods employing weighted voting across various preprocessing scenarios. By focusing on metrics like Hamming Loss and examining the impact of simplified preprocessing, this study aims to provide a balanced approach to preserving semantic richness while enhancing classification performance for Qur'an verses. The findings will contribute to bridging the gap between traditional and modern methodologies in multi-label classification.