

## Tense Identification in English Sentences Using an Ensemble Machine Learning Approach

IDTD Rathnayaka<sup>1</sup> and RAHM Rupasingha<sup>#</sup>

<sup>1</sup>Department of Economics and Statistics, Faculty of Social Sciences and Languages,  
Sabaragamuwa University of Sri Lanka

<sup>#</sup>hmrupasingha@gmail.com

For efficient communication, comprehension, and writing in a variety of contexts including academic papers, legal documents, and media documents proper use of tense is essential. Notwithstanding its significance, tense usage is a common error for both experts and students. By having access to a built-in tense detection tool, students can easily enhance their sentence construction skills through consistent self-study in early life education. The study objective was to propose an automatic classification of English sentences using machine learning algorithms according to their tense: past, present, or future. We used a dataset with 1500 sentences that were split equally across the three tense groups. Tokenization and lowercasing were used in the preprocessing phase, and the Term Frequency-Inverse Document Frequency (TF-IDF) technique was then used to extract features. Then, six classification algorithms Naive Bayes, Random Forest, Decision Tree (J48), Support Vector Machine (SVM), Logistic Regression and ensemble learning by combining the above five algorithms were tested. Metrics including accuracy, precision, recall, F-measure, and error values were used in the evaluation. In terms of evaluation, the ensemble learning strategy outperformed individual models in all evaluations by achieving the best accuracy of 95.56%. In ensemble learning, the majority voting combination rule worked best in 70% training data. This work shows how machine learning may improve tense classification, providing a useful tool for both academic and professional contexts.

**Keywords:** *tense identification, machine learning, ensemble learning, classification, English sentence*