

An Ensemble Machine Learning Approach for Categorizing English Sentences Based on their Emotional Tone

DBY Perera¹ and RAHM Rupasingha^{1#}

¹Department of Economics and Statistics, Faculty of Social Sciences and Languages
Sabaragamuwa University of Sri Lanka

[#]hmrupasingha@gmail.com

Communicating in English is a common thing in day-to-day life, since it is influential to understand the tone or the idea of English sentences. There will be misunderstandings regarding the attitude, sentiments, and point of view of the speaker or writer if we are unable to identify the appropriate sentence tone. Therefore, the objective of this study was to propose a machine learning based ensemble approach for identifying the tone of the English sentences automatically. Here we collected around 1000 English sentences as the input data. This approach used lowercasing, stopping words removal, lemmatization and tokenization as preprocessing techniques. Then Term Frequency-Inverse Document Frequency (TF-IDF) feature vector extraction method was used to extract features from the sentences. For classification five different individual ML algorithms namely Nave Bayes, Support Vector Machine (SVM), Logistic Regression, Random Forest, and Decision tree were applied. Then, an ensemble learning algorithm was applied by combining those five algorithms through the average probability method. The experimental results show that the ensemble learning algorithm outperforms the individual algorithms in terms of accuracy (96.9%), precision (97.2%), f-measure (96.9%), recall (96.9%) as well as Mean Absolute Error (MAE)(0.0113) and Root Mean Square Error (RMSE) (0.0163) for measuring error. The study introduced an ML based ensemble method, incorporating preprocessing techniques and TF-IDF feature extraction, which outperforms individual algorithms in classifying English sentences based on their tone.

Keywords: *machine learning, ensemble learning, sentence tone analysis, classification*