A Feature-Enriched Machine Learning Approach for Sarcasm Detection in Spanish Conversational Dialogues

Rafael Jimenez Castro[1]

[1]*Department of Electrical and Computer Engineering, Universidad Autónoma de Ciudad Juárez, Mexico*

Abstract. Sarcasm detection remains a challenging task in natural language processing, and particularly in short and informal text where contextual cues are limited to the text itself. This research proposes a machine learning model for sarcasm detection in Spanish dialogues by incorporating context from the conversation. A dataset was constructed from scripted dialogues, then enriched with contextual features such as sentiment polarity, topic modeling, and speaker information. BERT-based models were then used to generate semantic embeddings augmented with contextual features and reduced using feature selection, before being classified using a multilayer perceptron (MLP). The evaluation was conducted using cross-validation and achieved an F1 score of 0.9195. The proposed framework also improves sarcasm detection performance on the Sarcasm Headlines Dataset, achieving an F1-score of 0.9914, which represents a 3.05% improvement over previously published models.

Keywords: Sarcasm Detection, Contextual Sarcasm, Dialogue-Based NLP, Spanish NLP