ABSTRACT

GAN INDOBERT: GENERATIVE ADVERSARIAL LEARNING FOR EMOTION CLASSIFICATION WITH LIMITED LABELED DATA ON INDONESIAN TWEETS

By

Norman Muhammad Mikhail
19/444313/PA/19375

Performing emotion classification tasks on popular social media platforms such as Twitter using machine learning methods necessitates a vast amount of training data to achieve high accuracy. In many scenarios, particularly in the context of emotion classification tasks in the Indonesian language, procuring access to high-quality labeled data proves to be both expensive and time-consuming. One viable solution to this challenge is to prioritize the collection of unlabeled data, which is comparatively easier to amass. This paper proposes a semi-supervised learning approach to minimize the dependency on labeled data while still ensuring commendable performance in classification tasks. Specifically, we introduce the GAN IndoBERT model designed to classify emotions in Indonesian text. Our results highlight the model's superior performance across varying thresholds of labeled data availability. Notably, even with as little as 1% of labeled data, GAN IndoBERT achieved 47% accuracy, outstripping the traditional IndoBERT's performance by a substantial margin. Such findings underscore the potential of integrating generative adversarial networks with deep learning models, ensuring optimal performance in data-limited scenarios

Keywords: Emotion Classification, Semi-supervised Learning, Generative Adversarial Networks, GAN IndoBERT, Indonesian Text, Data Augmentation, Machine Learning, Social Media Analysis.