Data Augmentation using Multi-Turn Dialogue Prompting for Sentiment Analysis

Syifa Fatimah Azzahrah[1], Ade Romadhony[2]

[1,2]Fakultas Informatika, Universitas Telkom, Bandung
[1]syifafatimahh@students.telkomuniversity.ac.id,
[2]aderomadhony@telkomuniversity.ac.id

Abstract

Label imbalance and data scarcity in Natural Language Processing (NLP) pose significant challenges to the development of effective text classification models. One approach to solve label imbalance and data scarcity is data augmentation. In this research, we examine the impact of multi-turn dialogue prompting approach on a large pretrained language model based chatbot for data augmentation on sentiment analysis task. Model evaluation on original dataset before data augmentation was performed shows accuracy of 0.6 and an average F1 score of 0.57. This performance reflects non-uniformity of labels and poor performance on the original dataset. After data augmentation was performed, the model performance improved with an accuracy score of 0.9 and F1-score of 0.9. The performance increase shows that data augmentation can significantly improve classification accuracy, especially for imbalance datasets.

Keywords: *data augmentation, multi-turn dialogue, sentiment analysis*