

Main findings

Code Repository: https://github.com/anondo1969/MultiNERD_NER_Evaluation

Author: [Mahbub Ul Alam](#) (mahbub.ul.alam.anondo@gmail.com)

The experiments conducted on Models A and B using BERT for Named Entity Recognition (NER) reveal distinct insights and limitations. Model A's '[segeval](#)' results show satisfactory performance with a correlation between scores and the number of entity instances. Surprisingly, rare entities like INST achieved high scores, despite the model's fine-tuning on a small dataset. However, the use of a cased BERT model raises questions about the influence of casing on these results. In contrast, Model B, while improving overall, showed a slight decrease in scores for individual entities compared to Model A. This suggests that Model A's performance benefited from the diversity of entities. '[nervaluate](#)' results from both models indicate moderate satisfaction with 'partial' evaluation in Model A, while Model B demonstrated consistent scoring across various metrics. This consistency implies uniform performance across entities. The comparison between cased and uncased BERT models revealed that cased models unexpectedly outperformed uncased ones. This result, potentially influenced by limited hyperparameter selection, warrants further exploration. Limitations of the study include constrained computational resources, limiting extensive hyperparameter tuning and exploration of other transformer-based models. Future work aims to address these constraints and explore data distribution within the dataset for real-world applicability, guided by methodologies such as in Fu et al. ([source](#)).