Task 4

BioBert recognises "diseases" and "drugs" with greater precision and accuracy at the semantic level of human judgement.

In en\_ner\_bc5cdr\_md, the most common "disease" entity occurs 24,630 times as "pain", p.o. occurs 11,607 times as the most common "drug", and so on. This confidence level of en\_ner\_bc5cdr\_md is low. The most common "disease" entity as "pain" occurs 24,630 times, p.o as the most common "drug" occurs 11,607 times, etc. However, in BioBert, the number of extractions for the entities are more accurate.

BioBERT is more effective in entity recognition tasks in extensive biomedical texts compared to en\_ner\_bc5cdr\_md. This is attributed to the fact that BioBERT is pre-trained on a large-scale biomedical text corpus, providing it with a broader understanding of the biomedical domain. On the other hand, en\_ner\_bc5cdr\_md is only pre-trained on the BC5CDR dataset, limiting its exposure to a more narrow range of biomedical text data.