Review of: "Enhancing Student Writing Skills: Leveraging Transfer Learning and Fine-tuned Language Models for Automated Essay Structure Recognition"

Evelin Amorim[1]

1 University of Porto

The paper proposed to detect argumentative structures in English-written essays. The methodology employs two LLMs, namely, BigBird and Longformer. Then, the author compares the results of the two models. This work has several problems. Next, I will list only a few of them.

1) **The paper lack of a proper problem definition:** The authors present some results in the abstract, but the classes or objects to be identified are unknown. There is only the description of "key argumentative and rhetorical elements" as the tasks to be solved. This is insufficient to understand the task to be tackled in the work. At first, when the author states that the essay structure will be evaluated, it is possible to think about grades as outputs. However, f1 is an inadequate measure for this type of task. Regardless of the metric, it is possible to notice later in the paper that the task is to classify sentences as lead, position, claim, counterclaim, rebuttal, evidence, and concluding statement. Thus, it is not the essay structure that will be evaluated as stated; it is the classification of elements of an argument.

2) **There is an absence of a decent description of the corpus:** Since the beginning, the author does not provide a clear description of the corpus. For instance, in the abstract, there is a claim about the results without even mentioning any characterization of the dataset. The language of the dataset is also not mentioned.

3) **The contributions of the paper are unclear.** The contributions of the author are not explicit. The dataset employed in the experiments seems well-known, but there is no mention of the state of the art. So, what is the real contribution of this work?

4) **The lack of a baseline system.** There is no baseline system. Usually, in papers of this kind, a baseline system is presented. Sometimes, the baseline is state of the art, and then the work proposes a system that has some advantage compared to the baseline.