

Review of: "Sentiment Analysis of Opinions about Online Education in the Kurdistan Region of Iraq during COVID-19"

Edgar León-Sandoval¹

¹ Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM)

Potential competing interests: No potential competing interests to declare.

Regarding form and English grammar, this article is in need of re-writing. Mayor changes are expected in terms of English writing, next are but a few examples.

- "That caused the world to descend into chaos", This reads a little sensationalist.
- "and the developed dataset developed",
- "The dataset is available under CC0 1.0 Universal license at <https://github.com/KurdishBLARK/SentimentAnalysis/tree/main/OnlineEducation-during-COVID-19>.", this should be a reference or an appendix.
- "Based on the points mentioned, we devise our method to conduct this research as we describe in Sections 4.", mixing plural with singular.

And many others. Now, the comments regarding content. Table 2 includes rather old studies, the conclusions gathered from there should be described in prose rather than another bulleted list, and each element of the list requires additional explanation. For example, "Tweepy and other applications interacting with Twitter's API work better for tweet retrieval" Leaves us, the readers, with the question of what other applications were considered there, and work better than what in tweet retrieval. That is not including the reasoning on why that would be important in the work, for even if the API is awful, the abstract mentions a dataset of 511 tweets. For such a small dataset, even manual retrieval is feasible.

In the method, the authors mention an expectation of a small number of tweets to be collected but fail to mention the reason or impact of this decision. In their own literature review, other have used a dataset ranging from 1200 to 58000 elements, which is what one would expect from this work.

Its bad form to utilize a bulleted list to explain the results. For the conclusions, the authors mention gathering 711 tweets, and then filtering them for a total of 512. Neither number match those expressed elsewhere in the article. All of the conclusions mentioned remark on the performance of their four classifiers and two data splitting approaches for Kurdish (Sorani). While it would be nice to see more classification models, and more modern data splitting techniques, I fail to see the relationship between this and the aforementioned "educational efficiency" of a remote model, nor I see the relationship of either of them with the COVID-19 mandatory lockdown.

I would expect major changes for this paper before being considered for publication. And the authors should, for I believe their work is interesting and worthwhile reading, if these points are properly addressed.

