Dear Mr. Wilde,

We have received the reports from our advisors on your manuscript, "Evolutionary Dataset Optimisation: learning algorithm quality through evolution", which you submitted to Applied Intelligence.

Based on the advice received, the Editor feels that your manuscript could be reconsidered for publication should you be prepared to incorporate major revisions. When preparing your revised manuscript, you are asked to carefully consider the reviewer comments which are attached, and submit a list of responses to the comments. Your list of responses should be uploaded as a file in addition to your revised manuscript.

Please make sure to submit your editable source files (i. e. Word, TeX).

In order to submit your revised manuscript electronically, please access the journal's website.

Your username is: ********
If you forgot your password, you can click the 'Send Login Details' link on the EM Login page at https://www.editorialmanager.com/apin/

Please click "Author Login" to submit your revision.

We look forward to receiving your revised manuscript.

Best regards,
Springer Journals Editorial Office
Applied Intelligence

COMMENTS FOR THE AUTHOR:

Reviewer #1: This paper needs Major revisions:
- I think the subject and object of this paper is very ambiguous in the introduction even though authors well described the previous work, I hope the author should consider why this paper is necessary to the read.
-In this paper, a new method is proposed for learning how algorithms -perform. Classically, algorithms are compared on a finite number of existing (or newly simulated) benchmark data sets based on some fixed metrics. The algorithm(s) with the smallest value of this metric is chosen to be the 'best performing'.
-The language needs to be revised by native speaker research.
-make the ABSTRACT as a single paragraph and make sure do present your work in clearer.
- You should improve the ABSTRACT.(rewrite the abstract to reflect the main idea and it's results, without any not suitable details) in the ABSTRACT alongside with the obtained results (the results you got it and what is the situation of your results in comparison with other published methods). Mentioned to the benchmarks which have been used in this paper.
-in figure 1, there are question marks, you did not explain that. why?
- Add in the and of section 1, add a new paragraph that presents the organization of the paper.
-Unify the symbols such as Figure or Fig.
- The related works section is not provided. I suggest to increase the number of studies and add anew discussion there to show the advantage, disadvantage, and weakness of the studied works.

Authors should discuss the literature review more deep and clearly.

-Most references in the list of reference are very old, therefore, must update the reference list to contain articles related off at least five years and indexing in ISI and Scopus Database, in general update that list by the following reference related to predictions: 1-Abualigah, L. M. Q. (2019). Feature Selection and Enhanced Krill Herd Algorithm for Text Document Clustering. Studies in Computational Intelligence. 2-Abualigah, L. M. Q., & Hanandeh, E. S. (2015). Applying genetic algorithms to information retrieval using vector space model. International Journal of Computer Science, Engineering and Applications, 5(1), 19. 3-Abualigah, L. M., & Khader, A. T. (2017). Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering. The Journal of Supercomputing, 73(11), 4773-4795. 4-Abualigah, L. M., Khader, A. T., & Hanandeh, E. S. (2018). Hybrid clustering analysis using improved krill herd algorithm. Applied Intelligence. 5-Abualigah, L. M., Khader, A. T., & Hanandeh, E. S. (2018). A Combination of Objective Functions and Hybrid Krill Herd Algorithm for Text Document Clustering Analysis. Engineering Applications of Artificial Intelligence. 6-Abualigah, L. M., Khader, A. T., & Hanandeh, E. S. (2017). A new feature selection method to improve the document clustering using particle swarm optimization algorithm. Journal of Computational Science.

- You need to explain clearly your proposed methods epically for the proposed method.
- Add a new figure to show the general procedures of the proposed method
- For the experimental results, it will be good to present a statistical test in the comparison of the results with other published methods. This can help to support the claim on improved results obtained with the selection methods studied.
- What are the pros and cons of the proposed method? Please respond to this question in the article text.
- in this paper, as you claimed, benchmark datasets are proposed. This is not clear in the paper

Reviewer #2: This is a very nice paper, with an original concept and a well-written description throughout. It is definitely worthy of publication and I only have minor queries:
The key concept behind the work is the evolutionary generation of datasets in order to understand the limitations and capabilities of ML algorithms. Thus, for this to be effective, you need to demonstrate that your representation and genetic operators are capable of permitting evolution to generate any possible dataset - or at least the set of datasets that would adequately cover the sets corresponding to the ML algorithm under investigation. e.g. If K-means needs to be investigated with sets A,B,C,D,E (each of which might be a different class of distribution) but your evolutionary approach is only capable of generating sets A,C,E, then you may gain an incomplete or misleading view of the capabilities of the algorithm. It is a big ask perhaps to prove that your method can generate *any* data distribution (or is not overly biassed towards the generation of some compared to others because of the representation or operators) and I'm sure later versions would be able to improve in these areas, but this issue is so important for the approach to be viable, that I think you should provide some evidence in the paper that the method can produce adequate coverage.
The second concern is the way evolutionary algorithms tend to find the easiest way out - so how can you stop the EA from evolving the simplest, easiest datasets for each ML algorithm (or the most difficult datasets if you reverse the fitness), instead of exploring the range of possibilities? (One option is to move more into curiosity-driven search, perhaps?)

****

https://www.editorialmanager.com/apin/login.asp?a=r). Please contact the publication office if you have any questions.