

First peer review report (FES)

Polina Pilyugina, Alexey Voskoboinikov

April 23, 2021

1 Are project objectives clearly explained? If not write what was not clear. Provide your opinion on the problem.

Project description and problem statement sections contain detailed information on the project objectives. In our opinion, feature selection process is important for a variety of machine learning tasks and thus creation of unified tool would be beneficial for all users. Also we believe that creation of first unification library may serve as a baseline for further contributors and as a framework for other methods' implementation, which makes this project exceptionally usefull in the long run.

2 Is proposed baseline solution relevant to the problem? Could you recommend something to the team?

Proposed baseline solutions are clearly explained and advantages and disadvantages of each method are outlined explicitly. Maybe it would be useful to additionally add some naive baseline approaches from sklearn.feature_selection library such as SelectKBest, for example, to compare results with these methods additionally.

3 Is it clear from the report how the team is going to test / evaluate the results? (not only metrics but code and bugs). Could you recommend something here?

It is clearly explained how the evaluation procedure will be structured, in terms of both metric choice and dataset selection. Additionally, we find it wise of the team to add noise testing as a part of evaluation, as it allows to understand the performance of the method in more detail. We suggest the team to add naive approaches as well, to evaluate the performance in more detail. Additionally, it might be useful to add evaluation for scalability of the solution in case of datasets with huge amount of features.

4 Is it possible to guess next project development steps from the report and github repository? What a team is going to do next, provide with 3 next possible steps?

- Implementation of IHT and Simultaneous Feature and Feature Group Selection through Hard Thresholding methods, following the sklearn API
- Creation of synthetics dataset and examination of model performance on this dataset, as well as Boston Housing dataset
- Benchmarking, evaluation and code finalization

5 Would you recommend to improve the report and how? Or it is all good.

This report is very sufficient and covers the proposed project in a very detailed way. Overall structure is very good and allows to easily follow the project outline. It would be useful to add the section with challenges of the implementation process as a separate subsection.

6 Is project github repository easy to follow?

The gitbub repository includes main points from the report: description of baseline solution, both methods to implement, metrics and data. Also, basic structure of the project is already in place.

Overall, chosen structure of the github repo allows to find each part of the project and thus helps to follow the process and induces reusability.