

1 Rue Honoré d'Estienne d'Orves
91120 Palaiseau
+33(0)761104782

December 28, 2016

Geoff Holmes
Action editor of JMLR

Dear Sir,

Alongside with the submission 16-365 at the MLOSS of the JMLR, we would like to thank the reviewers for their pertinent suggestions in an effort to improve our software. Therefore, we addressed the issues raised by the reviewers by improving the manuscript, the documentation, and the software.

Below, we go into details regarding the changes which has been done:

Reviewer #1:

- *I believe that a real world example of this package improving generalization performance would make the paper more convincing.*

Due to the page limits, we did not added any real-world example in the manuscript itself. However, examples have been added in the documentation of the software. Additionally, we continuously aim at adding real-world case studies.

- *Typos / phrasing suggestions.*
We addressed the typos and adopted the phrasing suggested.

Reviewer #2:

- *A table could be added with the implemented methods grouped by type (over-sampling, under-sampling ...). The table could also include a categorization of the methods concerning the possibility of application only to two-class or multi-class problems. This could provide a better overview of the implemented methods.*

We simplified the description for each method and add the suggested table indicating on which type of problems — binary *vs.* multi-class — the methods can be used.

- *It would help users with real world problems when using the proposed toolbox. Researchers and developers would also benefit from this perspective.*
As previously mentioned, due to the page limits, we did not added any real-world example in the manuscript itself. However, examples have been added in the documentation of the software. Additionally, we continuously aim at adding real-world case studies.
- *There is an huge number of methods proposed to address the problem of imbalanced domains. The proposed toolbox includes several pre-processing methods and two algorithm level methods (ensembles). This is on the short side in terms of what is currently available.*

The implementation of additional methods is planned for the next release of the toolbox and is currently in the issue tracker on GitHub. We are striving to include these new methods within the next 6 months.

- *For instance, an example could be provided for studying the impact of using different parameters with a given strategy. The results could then be presented considering different performance assessment metrics. Given that the issue of performance assessment in imbalanced domains is complex, this example could also show how the user may assess the performance in a new metric that is not implemented in scikit-learn (for instance G-mean, dominance or Index of Balanced Accuracy(IBA)).*

We implemented new metrics available in the `metrics` modules of the toolbox. Sensitivity/specificity, geometric-mean, and index of balanced accuracy have been added. Additionally, we proposed a method to generate a classification report reporting the different metrics. Furthermore, examples presenting the metrics have been added to the documentation.

- *Another example could also be added considering a real world data set. This example could be built to answer the question: “What is the best strategy that I can use in this specific problem?”*

We provide an example using a real-world dataset which illustrate the empirical search for such purpose. A grid-search using different balancing method in conjunction with a classifier is performed.

- *Finally, a third example that considers a problem with multiple classes could also be included.*

We introduce an example which illustrate the use of a method for a multi-class problem and we additionally generate the according classification report.

- *With respect to the references to implementations available in R, the authors should eventually refer the package UBL that is a specific R package for handling imbalanced distributions, so it is probably the R package with more similar objectives to the current one.*

We refer to this toolbox in the manuscript.

Thank you for your time and consideration.

We look forward to your reply.

Sincerely,

G. Lemaitre, F. Nogueira, and C.
Aridas