

# ClusterValidityIndices.jl: Batch and Incremental Metrics for Unsupervised Learning

Sasha Petrenko<sup>1</sup>

<sup>1</sup> Missouri University of Science and Technology

DOI: [10.21105/joss.03527](https://doi.org/10.21105/joss.03527)

## Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Adi Singh](#) ↗

## Reviewers:

- [@rMassimiliano](#)
- [@malmaud](#)

Submitted: 28 May 2021

Published: 23 July 2021

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## Summary

ClusterValidityIndices.jl is a Julia package for evaluating the performance the performance of clustering algorithms without the aid of supervised labels. Cluster Validity Indices (CVI) provide a metric of the over- or under-partitioning of an arbitrary clustering algorithm with only the original data and labels assigned by the clustering algorithm. Furthermore, there exist formulations of every CVI such that they may run incrementally (i.e. Incremental CVIs, or ICVI), streaming alongside the clustering algorithm and producing the same results as in their batch implementations. Using a standard interface, each CVI in this package can be run with any clustering algorithm to produce a metric of that algorithm's performance in scenarios where explicit supervised labels do not exist, which is extremely useful in real-world applications where that is often the case.

## Statement of need

CVIs are useful as one of the only methods of determining the performance of a clustering algorithm in the absence of explicit labels. Furthermore, ICVIs can measure the performance of clustering algorithms as they are running in a computationally tractable manner, which is incredibly useful in a variety of streaming clustering applications ([Brito Da Silva et al., 2020](#)).

There exist many CVIs in the literature, and their algorithmic and programmatic requirements are often very similar. Despite their utility in machine learning applications, however, there does not exist to date a unified repository of their implementations in Julia. Furthermore, new incremental variations of these algorithms are regularly developed in the literature without the ability to update the original implementations. The purpose of this package is to create a unified framework and repository of CVIs so as to fill the gap left by most metrics in this machine learning problem subset. # Acknowledgements

This package is developed and maintained with sponsorship by the Applied Computational Intelligence Laboratory (ACIL) of the Missouri University of Science and Technology. This project is supported by grants from the Army Research Labs Night Vision Electronic Sensors Directorate (NVESD), the DARPA Lifelong Learning Machines (L2M) program, Teledyne Technologies, and the National Science Foundation. The material, findings, and conclusions here do not necessarily reflect the views of these entities.

## References

- Brito Da Silva, L. E., Melton, N. M., & Wunsch, D. C. (2020). Incremental Cluster Validity Indices for Online Learning of Hard Partitions: Extensions and Comparative Study. *IEEE Access*, 8(i), 22025–22047. <https://doi.org/10.1109/ACCESS.2020.2969849>