

# Applying Semi-Automated Hyperparameter Tuning for Clustering Algorithms

Elizabeth Forest<sup>\*†</sup>, Anne Swinbourne<sup>†</sup>, Trina Myers<sup>‡</sup>, and Mitchell Scovell<sup>†</sup>  
<sup>†</sup> James Cook University, Townsville, Australia  
<sup>‡</sup> Queensland University of Technology, Brisbane, Australia  
<sup>\*</sup> Email: [elizabeth.forest@my.jcu.edu.au](mailto:elizabeth.forest@my.jcu.edu.au)

This preliminary study proposes a semi-automated framework for the hyperparameter tuning of clustering algorithms, using a grid search to develop a series of graphs and easy to interpret metrics that can then be used for more efficient domain-specific evaluation.

## Background:

When approaching a clustering problem, choosing the right clustering algorithm and parameters is essential, as each clustering algorithm is proficient at finding clusters of a particular nature [1]. Due to the unsupervised nature of clustering algorithms, there are no ground truth values available for empirical evaluation, which makes automation of the parameter selection process through hyperparameter tuning difficult [1]–[3]. Previous approaches to hyperparameter tuning for clustering algorithms have relied on internal metrics, which are often biased towards certain algorithms, or having some ground truth labels available, moving the problem into the semi-supervised space [2]–[4].

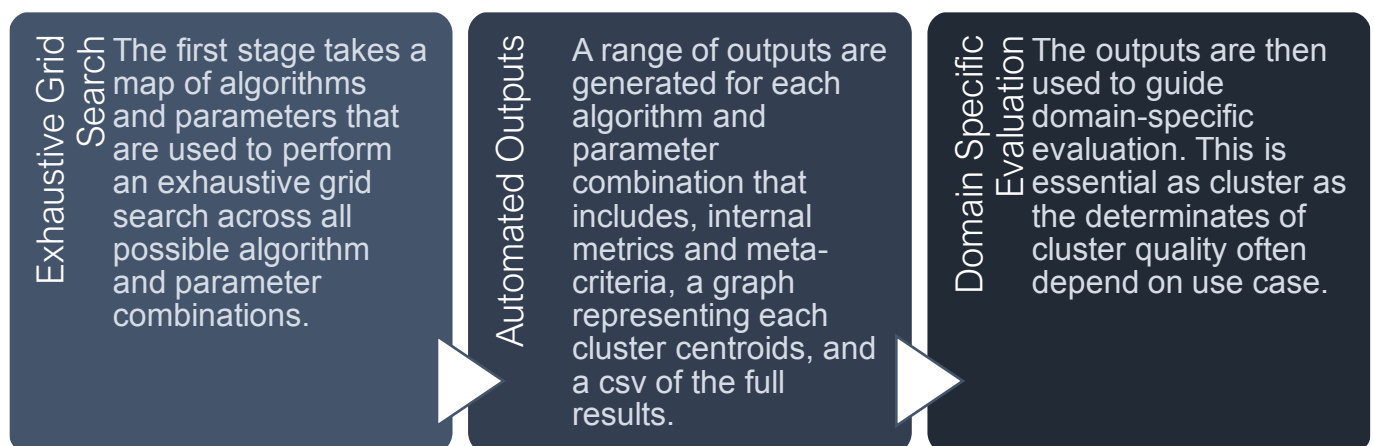
## Preliminary Results:

The framework was applied to compare three clustering algorithms and a variety of parameters for persona development. A total of 16 algorithm and parameter combinations were considered by the framework. Of the 16:

- 6 were quickly ruled out by the internal metrics and meta criteria
- 4 were ruled out by the graphs
- Leaving 6 for domain-specific evaluation

The algorithm and parameter combination found most appropriate during the domain-specific evaluation contradicted the results of a completely automated method based on internal metrics.

## The Framework:



## Conclusion:

The preliminary results show that a semi-automated hyperparameter tuning framework for clustering algorithms facilitates efficient domain-specific evaluation that can produce results relevant to the given problem area.

- [1] A. K. Jain, 'Data clustering: 50 years beyond K-means', *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, Jun. 2010, doi: 10.1016/j.patrec.2009.09.011.
- [2] X. Fan, Y. Yue, P. Sarkar, and Y. X. R. Wang, 'On hyperparameter tuning in general clustering problems', in *Proceedings of the 37th International Conference on Machine Learning*, Jul. 2020, vol. 119, pp. 2996–3007. [Online]. Available: <http://proceedings.mlr.press/v119/fan20b.html>
- [3] T. Van Craenendonck and H. Blockeel, 'Constraint-based clustering selection', *Mach Learn*, vol. 106, no. 9, pp. 1497–1521, Oct. 2017, doi: 10.1007/s10994-017-5643-7.
- [4] U. Von Luxburg, R. C. Williamson, and I. Guyon, 'Clustering: Science or art?', 2012, pp. 65–79.